










RMD
OpenRheumatic &
Musculoskeletal
Diseases

ORIGINAL RESEARCH

Natural language processing to identify and characterize spondyloarthritis in clinical practice

Diego Benavent ^{1,2}, María Benavent-Núñez ^{1,3}, Judith Marin-Corral ¹, Javier Arias-Manjón ⁴, Victoria Navarro-Compán ⁵, Miren Taberna ¹, Ignacio Salcedo,⁴ Diana Peiteado ⁵, Loreto Carmona ⁶, Eugenio de Miguel ⁵, Savana Research group¹

To cite: Benavent D, Benavent-Núñez M, Marin-Corral J, *et al.* Natural language processing to identify and characterize spondyloarthritis in clinical practice. *RMD Open* 2024;**10**:e004302. doi:10.1136/rmdopen-2024-004302

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/rmdopen-2024-004302>).

Received 5 March 2024
Accepted 7 May 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Savana Research S.L, Madrid, Spain

²Rheumatology, Hospital Universitari de Bellvitge,

L'Hospitalet de Llobregat, Spain

³Nutrition Department, CEU San Pablo Monteprincipe School, Madrid, Spain

⁴Medsavana S.L, Madrid, Spain

⁵Rheumatology, Hospital Universitario La Paz, Madrid, Spain

⁶Instituto de Salud Musculoesquelética, Madrid, Spain

Correspondence to

Dr Diego Benavent;
d_benavent@hotmail.com

ABSTRACT

Objective This study aims to use a novel technology based on natural language processing (NLP) to extract clinical information from electronic health records (EHRs) to characterise the clinical profile of patients diagnosed with spondyloarthritis (SpA) at a large-scale hospital.

Methods An observational, retrospective analysis was conducted on EHR data from all patients with SpA (including psoriatic arthritis (PsA)) at Hospital Universitario La Paz, between 2020 and 2022. Data were collected using Savana Manager, an NLP-based system, enabling the extraction of information from unstructured, free-text EHRs. Variables analysed included demographic data, SpA subtypes, comorbidities and treatments. The performance of the technology in detecting SpA clinical entities was evaluated through precision, recall and F-1 score metrics.

Results From a hospital population of 639 474 patients, 4337 (0.7%) patients had a diagnosis of SpA or their subtypes in their EHR. The population predominantly comprised men (55.3%) with a mean age of 50.9 years. Peripheral SpA (including PsA) was reported in 31.6%, axial SpA in 20.9%, both axial and peripheral SpA in 3.7%, while 43.7% of patients did not have the SpA subtype reported. Common comorbidities included hypertension (25.0%), dyslipidaemia (22.2%) and diabetes mellitus (15.5%). The use of conventional disease-modifying antirheumatic drugs (csDMARDs) and biological DMARDs (bDMARDs) was documented, with methotrexate (25.3% of patients) being the most used csDMARDs and adalimumab (10.6% of patients) the most used bDMARD. The NLP technology demonstrated high precision and recall, with all the assessed F-1 score values over 0.80, indicating reliable data extraction.

Conclusion The application of NLP technology facilitated the characterisation of the SpA patient profile, including demographics, clinical features, comorbidities and treatments. This study supports the utility of NLP in enhancing the understanding of SpA and suggests its potential for improving patient management by extracting meaningful information from unstructured EHR data.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Spondyloarthritis (SpA) involves a spectrum of inflammatory diseases with both axial and peripheral manifestations, including conditions such as psoriatic arthritis.
- ⇒ Despite advancements in understanding the epidemiology and clinical features of SpA, challenges remain in accurately characterising and diagnosing these conditions.
- ⇒ The potential of artificial intelligence (AI), specifically natural language processing (NLP), in healthcare research is recognised, yet its application in studying SpA through electronic health records (EHRs) is limited.

INTRODUCTION

The spectrum of inflammatory manifestations captured in patients with spondyloarthritis (SpA) has drawn increasing interest in recent years. SpA encompasses mainly two main disease subtypes characterised by shared genetic factors and clinical presentations, with axial (axSpA) or peripheral (pSpA) predominant manifestations, frequently associated with enthesitis, dactylitis and extramusculoskeletal manifestations (EMMs), such as uveitis, inflammatory bowel disease or psoriasis.^{1,2} Besides, a particular type of presentation known as psoriatic arthritis (PsA) is often associated to the SpA spectrum.³ Despite significant progress in the fields of epidemiology, clinical features, imaging and biomarkers for the different subtypes of SpA, controversies persist in their characterisation and diagnosis frames.^{4,5}

The implementation of emerging methodologies, such as large-scale data analysis and artificial intelligence (AI), holds promise to enhance our understanding of the diverse clinical manifestations of diseases, and even to facilitate the identification of biomarkers.⁶ The use of electronic health records (EHRs)

WHAT THIS STUDY ADDS

- ⇒ This study highlights the application of NLP in the characterisation of the SpA patient profile, including demographics, clinical features, comorbidities and treatments, from unstructured EHR data.
- ⇒ Among 4337 patients identified with SpA, detailed analyses of disease subtypes, peripheral manifestations, extramusculoskeletal manifestations and prescribed treatments are provided, demonstrating the utility of NLP in extracting and interpreting complex clinical information.
- ⇒ The performance of the NLP methodology in accurately identifying SpA-related clinical entities is validated with high precision, recall and F-1 scores, showing the reliability of NLP in clinical data extraction.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ The findings encourage the integration of NLP and other AI methodologies in rheumatology research, providing a new avenue for detailed patient characterisation and disease understanding without the limitations of structured data.
- ⇒ Clinicians and healthcare providers might leverage insights from such analyses to tailor patient management strategies more effectively, acknowledging the diverse clinical presentations within the SpA spectrum.
- ⇒ Researchers could be encouraged to adopt NLP-driven approaches in the evaluation of real-world data, facilitating the generation of evidence-based guidelines and improving patient care in SpA.

has led to the gathering of substantial amounts of electronically stored information, enabling data extraction by using sophisticated techniques that would not be feasible to perform manually. However, despite evolving research in the field, the analysis of real-world clinical data automatically extracted from EHRs in patients with SpA remains scarce.⁷ Natural language processing (NLP), a branch of AI that merges linguistics with machine learning and deep learning models, is becoming a critical tool in this transformation.⁸ NLP involves the application of algorithms to identify and extract natural language, enabling computers to understand, interpret and generate human language.⁸ NLP may be instrumental in extracting and interpreting meaningful information from unstructured data sources, helping in computer-aided processing and analysis of human language.

Studies aiming to retrieve information of rheumatic diseases through NLP are scarce. Nonetheless, analyses of the free text to identify patients with SpA have already been performed in pilot studies.^{9 10} A recent study compared the usefulness for identification of algorithms using NLP versus International Classification of Diseases (ICD) codes. NLP significantly outperformed traditional approaches in identifying patients with axSpA, achieving a sensitivity of 0.78, a specificity of 0.94 and an area under the curve (AUC) of 0.93, compared with an AUC of 0.80–0.87 for the ICD-based methods alone.⁹ Evidence on the analysis of detailed characteristics of patients, beyond just disease identification, is limited.

To confirm the utility of NLP for research in the complex field of SpA, we characterised the patients with SpA diagnosed at our centre by extracting clinical information from unstructured free text. More concretely, we described their clinical features, associated comorbidities and treatment using a tool based on NLP.

METHODS

Study design and participants

This study was an observational, retrospective, non-interventional investigation focused on the secondary use of clinical data from EHR of all patients with a diagnosis of SpA (including PsA) in their EHR at the Hospital Universitario La Paz, within the years 2020 and 2022.

Data source and collection

The patient data used in this study came from diverse EHR sources, including outpatient consultations, inpatient wards and the emergency department. Savana Manager, an AI-based data extraction system using EHRead technology (Medsavana, Madrid, Spain), was used for data collection.¹¹ This tool facilitates using unstructured, free-text information from EHRs in research, ensuring patient anonymity and employing computational linguistic techniques for clinical context detection and validation.^{12 13} EHRead Technology is a complex clinical NLP pipeline combining a rich set of NLP techniques in a big data processing pipeline and has been successfully applied in a wide range of real-world evidence studies.^{14 15} This technology can extract vital details, collecting important clinical aspects, negations or time-related factors, thus allowing the synthesis of a patient database (online supplemental text 1). The terminology employed by Savana Manager is founded on the principles of Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) and contains an extensive array of over 1 000 000 descriptions, including medical concepts, abbreviations and laboratory parameters.^{16 17} Of note, data such as date of birth, age, sex and document dates are received from the hospital in a structured format. Structured data are organised in relational spreadsheets, where it is organised into tables, rows and columns, allowing for efficient processing and analysis—and does not need to be processed with the NLP pipeline as the rest of the data.

Data collected included a wide range of variables to assess the profile of patients with SpA. These variables included demographic data (gender, age), length of stay, SpA types (axial, peripheral), family history related to SpA, peripheral manifestations (dactylitis, enthesitis), EMMs (inflammatory bowel disease, psoriasis and uveitis), laboratory data (C reactive protein), number of radiographic exams performed (lumbar spine, and hands), comorbidities and treatments (non-steroidal anti-inflammatory drugs (NSAIDs), corticosteroids, conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), and biological DMARDs (bDMARDs)—adalimumab,

infliximab, etanercept), which were not necessarily prescribed for SpA manifestations. Data collection was completed on 22 September 2023, and data between 1 January 2020, and 21 October 2022 (*study period*) were included. The inclusion date was defined as the first date on which patients were detected in the study period. The follow-up period, in which all variables were searched, comprised all the records from the inclusion date until the last record for each patient.

On identification of patients with SpA, terminological entities found in their EHRs were systematically classified into sections, including demographics, medical history, presenting complaint, prescribed medications and diagnoses, among others. The SpA diagnoses were established based on the information jotted down by healthcare professionals (specialists or allied health workers) in the hospital EHR and included all linguistic hyponyms and synonyms in SNOMED terminology (online supplemental table 1).

Evaluation of data extraction

The purpose of this external validation process was to evaluate the performance of EHRead technology in detecting the SpA clinical entities or related variables.¹¹ Following Savana's validation methodology and using Savana's SampLe Calculator for the Evaluation (SLiCE), which calculates the minimum quantity of annotated EHRs needed to achieve desired parameters,¹¹ a random collection (corpus) of clinical documents was created and was subsequently manually annotated by three clinical experts.

The SLiCE tool established the range of 125–150 documents as optimal for this evaluation. After selecting 150 documents, the collection of the final corpus of documents resulted from the exclusion of documents that would not contribute meaningful information for annotation ($n=7$), yielding a total of 143 documents. From each selected document, the clinicians simultaneously annotate the variables defined; in case of any discrepancies or for validation of the consensus decision, a third clinician intervenes. The 'annotation gold standard' was defined for each annotated variable. Ten variables were selected for validation, considering their relevance for the study and the limited resources for data annotation. This selection was informed by an expert rheumatologist, aiming to involve population definition, disease features and treatment use.

Using this as the benchmark, we evaluated the EHRead performance by calculating precision, recall and F-1 score, standard metrics used to determine the reliability of the data extraction.

Precision, reflecting the reliability of the data extracted by the system, was determined using the formula $p=tp/(tp+fp)$. Recall, measuring the system's ability to capture relevant information, was computed as $R=tp/(tp+fn)$. The F-1 score, representing a balance between precision and recall, was derived using the equation $F=2 \times \text{precision} \times \text{recall}/(\text{precision}+\text{recall})$, serving as a measure of

the system's overall efficiency in information retrieval. For these calculations, true positives (tp) represented the total number of accurately identified entries, false negatives (fn) represented the instances of missed entries and false positives (fp) accounted for the entries that were incorrectly identified by the system.

Statistical analysis

The total number of patients with a diagnosis of SpA that visited the hospital during the study period was calculated. The comorbidities among patients with SpA were established by the total number of patients diagnosed with the specific condition during their follow-up. The frequency of past or current clinical characteristics and treatments was calculated. Qualitative variables were represented as absolute frequency and percentage, and quantitative variables as mean and SD.

To explore the reliability of the tool in different periods, we divided the study period into the two distinct time-windows—from 1 January 2020 to 30 April 2021 (period 1), and from 1 May 2021 to 21 August 2022 (period 2) and compared the results between relevant variables. For each metric compared across the two periods, Z values were computed and interpreted against critical values to assess significance; a Z-statistic with an absolute value greater than 1.96 indicated a statistically significant difference between the proportions at a 95% CI.

Ethics

The study received approval from the Hospital Universitario La Paz Independent Ethics Committee (Code PI-5619). Data processing and pseudonymisation were handled by the hospital information service before the transmission of data to Medsavana, ensuring that identifiable data were never received. Thus, the results only pertain to aggregated data, and no identification of patients or physicians was possible, thus fully complying with the European Union General Data Protection Regulation. This approach was carefully considered and submitted for evaluation by the corresponding ethics committee, who assessed the societal value of the research objectives, the impracticality of obtaining informed consent due to the pseudonymised data format and the secondary use of the EHR's in which there would be no interaction with any of the participant patients. Further information on the pseudonymisation process can be found in online supplemental text 2.

The study was conducted in accordance with legal and regulatory requirements and followed generally accepted research practices described in the Helsinki Declaration and applicable local regulations.¹⁸

RESULTS

Out of the 639 474 patients who attended the Hospital Universitario La Paz during the study period, 4337 had a SpA diagnosis reported within their EHR, which represents about 0.68% of the total hospital population. Particularly, 1373 (31.6%) were reported as having

Table 1 Demographic and clinical characteristics

	SpA N=4337
Demographics and toxic habits	
Sex (male)	2397 (55.3%)
Age	50.9 (12)
Active smoker	871 (20.1%)
Family history of SpA	
Spondyloarthritis	92 (2.1%)
Inflammatory bowel disease	59 (1.4%)
Psoriasis	199 (4.6%)
Uveitis	13 (0.3%)
Type of involvement	
Peripheral spondyloarthritis*	1373 (31.7%)
Axial spondyloarthritis	908 (20.9%)
Axial and peripheral spondyloarthritis	162 (3.7%)
Non-specified type	1894 (43.7%)
Extramusculoskeletal manifestations	
Uveitis	1308 (30.2%)
Inflammatory bowel disease	441 (10.2%)
Psoriasis	1251 (28.8%)
Peripheral manifestations	
Enthesitis	887 (20.5%)
Dactylitis	323 (7.5%)
Variables are expressed as n (%) or mean (SD). *Includes psoriatic arthritis. SpA, spondyloarthritis.	

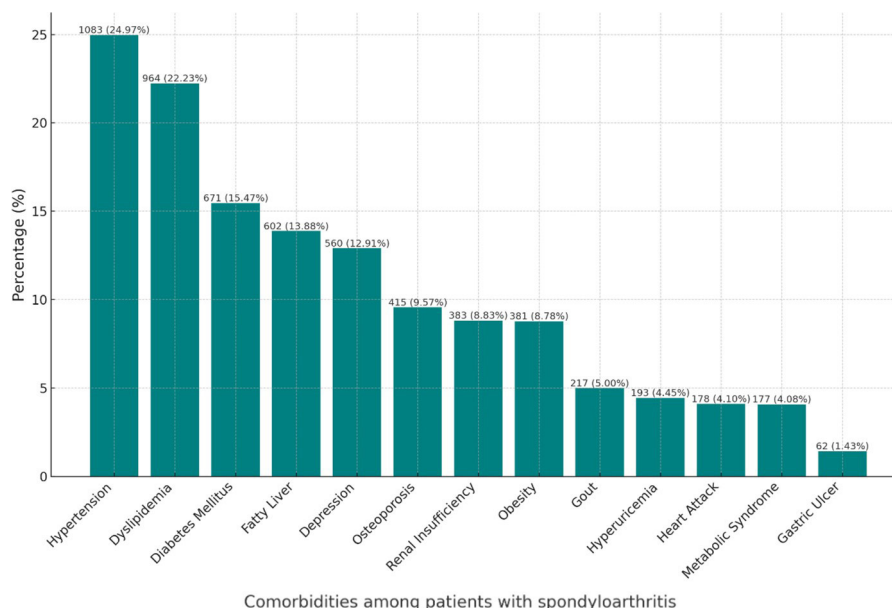
peripheral SpA (including pSpA and PsA), 908 (20.9%) as axSpA and 162 (3.74%) as axial and peripheral SpA, while the rest of patients (43.7%) did not have the SpA subtype reported.

Among all patients with SpA, 55.3% were men, their mean age (SD) was 50.9¹⁵ years and 20.0% were active smokers. Ninety-two patients (2.1%) had a reported family history of SpA as directly reported in their EHR (table 1). Concerning peripheral manifestations, 887 (20.5%) had enthesitis, and 323 (7.4%) had dactylitis, while EMMs were more commonly reported, including 30.2% patients with uveitis, 28.8% with psoriasis and 10.1% with inflammatory bowel disease.

Regarding diagnostic tests, 2195 (50.6%) patients had at least one elevated C reactive protein reported in their EHR during the study period. Moreover, 24.5%, 27.3% and 14.5% of patients with SpA had undergone pelvic, lumbar spine and hands radiographic exams during the study period, respectively.

In terms of reported comorbidities, cardiovascular risk factors were prominent, with 1083 (25.0%) patients presenting hypertension, 964 (22.2%) dyslipidaemia and 671 (15.5%) diabetes mellitus. Besides, 602 (13.9%) presented fatty liver, 560 (12.9%) suffered from depression, 415 (9.6%) had osteoporosis and 381 (8.8%) had obesity reported. All the assessed comorbidities can be seen in figure 1.

As for treatment prescribed, the use of NSAIDs included 26.7% of patients taking dexketoprofen, 24.4% taking ibuprofen, 21.9% etoricoxib, 17.6% naproxen and 6.2% taking celecoxib (figure 2). Concerning corticosteroids, 23.5% used prednisone, 15.1% dexamethasone, 8.3% prednisolone and 1.8% deflazacort. Concerning

**Figure 1** Comorbidities detected in patients with spondyloarthritis.

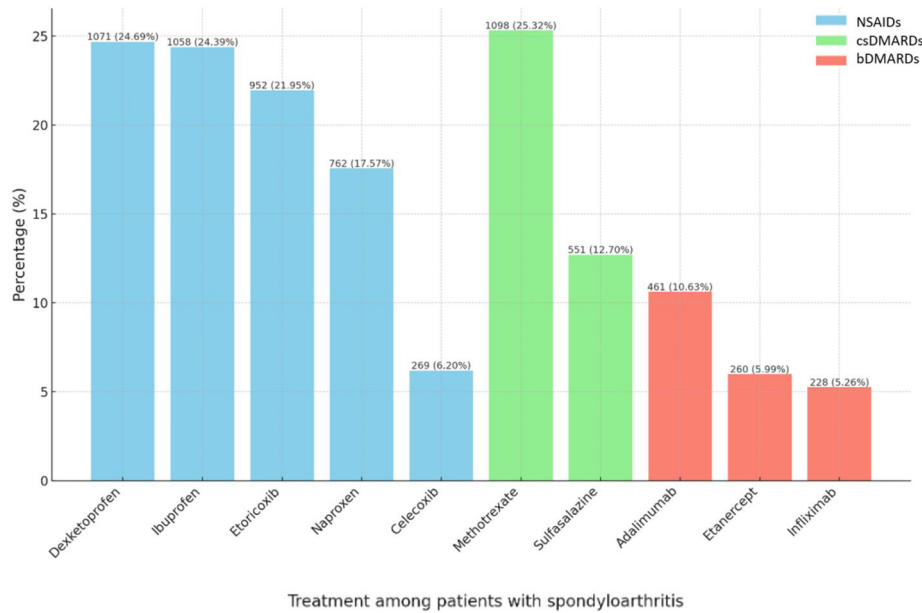


Figure 2 Treatments detected in the spondyloarthritis population. bDMARD, biologic disease-modifying antirheumatic drugs; csDMARD, conventional synthetic disease-modifying antirheumatic drugs; NSAIDs, non-steroidal anti-inflammatory drugs.

csDMARDs, methotrexate was the most frequently used (25.3%) followed by sulfasalazine (12.7%) and then bDMARDs, with 10.6% on adalimumab, 6.0% on etanercept and 5.3% on infliximab.

In the analysis comparing two time windows, we assessed 20 variables across demographics, toxic habits, type of involvement, EMMs, peripheral manifestations, cardiovascular risk factors and treatments. The results indicated no statistically significant differences for the majority of these variables, demonstrating stability across the examined factors (online supplemental table 2). The only exceptions were found in the frequencies of sex and uveitis, which showed significant changes between the periods.

Information on the performance of the data extraction can be found in table 2. The linguistic evaluation of the SpA and axSpA variables yielded precision, recall and F-1 scores

of over 0.9, while the remaining variables demonstrated an F-1 score of over 0.80, thus reassuring the data presented.

DISCUSSION

Our study assessed the clinical profile of patients with SpA in a large-scale hospital population, including a description of demographics, clinical characteristics, comorbidities and treatments using a novel technology, the Savana Manager NLP system. In this work, we apply this methodology to SpA. Our findings highlight the reliability of this technology, in accurately identifying and characterising variables as evaluated by the performance evaluation. While there are already some studies that benefit from the collaboration of expert technologists and NLP-specialised clinicians,^{14 15} our approach involved using a

Table 2 Performance metrics of detection of the assessed variables

	TP	FP	FN	Annotations	Precision	Recall	F1-score
Uveitis	277	1	1	278	0.996	0.996	0.996
Dactylitis	94	3	2	96	0.969	0.979	0.974
SpA	147	5	3	150	0.967	0.980	0.974
Enthesitis	392	2	38	430	0.995	0.912	0.951
AxSpA	16	1	1	17	0.941	0.941	0.941
AS	73	20	4	77	0.785	0.948	0.859
pSpA	4	0	2	6	1.000	0.667	0.800
Infliximab	11	0	0	11	1.000	1.000	1.000
Etanercept	32	1	1	33	0.970	0.970	0.970
Adalimumab	46	3	0	46	0.939	1.000	0.968

AS, ankylosing spondylitis (currently r-axSpA); axSpA, axial SpA; FN, false negative; FP, false positive; pSpA, peripheral SpA; SpA, spondyloarthritis; TP, true positive.

fit-for-purpose tool to extract clinical information from unstructured free text in EHRs. This tool can enhance our understanding of SpA features and patient management in real life, opening a new path for assessing the primary disease and its associated conditions. Compared with the classic model of collecting data from medical records, which limits the sample size due to the time necessary for data extraction, new NLP-based methodologies represent a qualitative leap. They transition from the laborious collection of data on tens or hundreds of patients to the automated collection of thousands of patients, including as many variables as deemed necessary.

Analysing our results critically, we found that the prevalence of SpA in our hospital population was approximately 0.7%, aligning with the reported prevalence rates in global population-based studies.^{19 20} Our study population showed a male predominance, consistent with prior studies,^{21 22} although the gender ratio is slightly lower than typically reported in SpA studies and close to gender parity. This might be affected by the inclusion of the entire spectrum of SpA in this study, as PsA and pSpA have shown a women predominance across studies.²³ This may also be influenced by the increased diagnosis of non-radiographic forms in women following the introduction of the Assessment of Spondyloarthritis International Society (ASAS) classification criteria.²⁴

Concerning peripheral manifestations, enthesitis was detected in 1 in 5 patients, while dactylitis was observed in less than 1 in 10. These findings are clinically reasonable considering the variability in reported prevalence rates across existing literature. For instance, the multinational ASAS-perSpA study documented enthesitis ever in approximately half of participants.²⁵ However, when stringent criteria for confirmation through specific diagnostic investigations were applied, a prevalence range of 13%–26% was reported, more closely aligning with our observations. It is possible that clinicians in our study were more inclined to report enthesitis in the EHR only when backed by a higher degree of diagnostic certainty. The prevalence of dactylitis, on the other hand, demonstrates considerable variation contingent on the specific subtype of SpA, with the ASAS-perSpA study indicating a prevalence of 6% in axSpA and 37% in PsA, reflecting the entirety of a patient's medical history. Conversely, our findings also unveiled a higher incidence of uveitis, with more than one-fourth of patients reporting an episode, in contrast to a maximum of one-fifth in the axSpA population noted in the ASAS-perSpA study.²⁵ This elevated occurrence may be attributable to data collection involving all consultations, including those of the ophthalmology department. It is crucial to underscore that, while the scope of our study was confined to data explicitly documented within a 3-year span in the medical records, all departments in the hospital were assessed, potentially influencing some results as compared with other means of collecting data.

Data from our study illustrate the substantial prevalence of comorbidities in patients with SpA, especially

cardiovascular risk factors. Indeed, a striking aspect was the reported prevalence rates of active smoking, hypertension and diabetes among patients with SpA. In this sense, the ASAS-COMORbidities in SPondyloArthritis (COMOSPA) multinational study had previously offered detailed insights into comorbidities among patients with SpA, already indicating an elevated prevalence of cardiovascular factors in a cohort of patients with relatively similar age (51 years in our cohort vs 45 in ASAS-COMOSPA).²⁶ One out of four patients with SpA in that study presented hypertension, particularly men in specific age groups, with a standardised risk ratio of 1.5 compared with non-SpA population. Hypercholesterolaemia rates in ASAS-COMOSPA were similar to the detected dyslipidaemia in our study, in about one-fifth of patients. This points to the necessity for screening and addressing these risk factors in this population, thereby potentially reducing the elevated cardiovascular risk and contributing to improved overall health outcomes. The concordance between our study's findings and those reported in ASAS-COMOSPA supports the reliability of novel NLP methodologies for automated data extraction. These AI-driven processes facilitate the rapid extraction of numerous study variables, contrasting with the potential time-consuming and resource-intensive data collection undertaken by researchers across various centres in traditional studies. Integrating AI for process automation is essential for the swift acquisition of results in clinical practice, enhancing efficiency and resource allocation.

The considerable prevalence of other various comorbidities, including depression, obesity, gout, renal failure and osteoporosis among the patients with SpA detected in our assessment also needs attention. It has been recently reported that patients with two or more comorbidities have more severe symptoms and worse functional status at baseline and over 2 years.²⁷ Additionally, these patients are more likely to stop their first TNF inhibitor (38.2%) than those with fewer comorbidities (26.6% and 25.4% for those with 0 and 1 comorbidity, respectively). The presence of these conditions could influence the choice of treatment, disease prognosis and the quality of life of patients, and therefore the importance of having a clear picture of their frequency in real-world patients is remarkable.

Our study also explored the use of various treatment modalities, from NSAIDs to csDMARDs and bDMARDs. NSAIDs were widely used in our population, with the most used drugs being dexamethasone and ibuprofen. csDMARDs, such as methotrexate, and bDMARDs, such as adalimumab were also frequently used. NSAIDs are the first-line treatment for patients with symptomatic axSpA.²⁸ Nonetheless, finding studies that specifically report the rate of real-world use of NSAIDs in SpA is challenging since most of the available literature focuses on the efficacy, safety and potential disease-modifying effects of NSAIDs.²⁹ The same holds true for the use of csDMARDs and bDMARDs in real-world studies, and

therefore the value of the reported use of these drugs in clinical practice is visible.

Our study presents strengths and weaknesses. Among the strengths, we include the large sample size of patients, including all individuals with SpA who attended the Hospital Universitario La Paz during the study period of almost 3 years, and the use of novel NLP techniques to extract data from EHRs. Moreover, these techniques were applied through an in-hospital application managed by the healthcare professionals, and an external validation was performed in a set of different variable categories and in two different time windows, ensuring the reliability and comprehensiveness of our findings. The linguistic evaluation of the SpA-related variables yielded a high precision, recall and F-1 score, which indicate that SpA diagnoses were accurately detected in the study population; most of the noted variables had an F-1 score greater than 0.85, which also points towards good reliability of the system. However, we acknowledge several limitations. As this is a single-centre study, the generalisability of our findings might be limited and further multicentre studies in real world using NLP are needed. An external validation of the algorithm is necessary to confirm its performance in different settings. Besides, the nature of the data extraction tool presents some limitations, such as the partial ability to stratify the study population into more specific subgroups once data are extracted or the possibility to design-specific NLP models tailored for the study. Indeed, data extraction was not performed according to specific patient subtypes (such as SpA and PsA). Moreover, for certain variables used in the study, such as some comorbidities or treatments, standardised metrics were not assessed due to feasibility reasons. Besides, treatments were detected as jotted down in the EHR of patients, and it may not have been prescribed for SpA; probably, dexamethasone was over-estimated due to the treatment of other diseases, such as COVID-19 (intravenous) or uveitis (eye drops). Thus, the findings related to these aspects should be interpreted with caution. Another limitation is that variables having some linked data, such as measurement instruments or HLA-B27, presented important limitations for their detection and were therefore not assessed. Despite acknowledging certain limitations of the Savana Manager, it is pertinent to emphasise its ongoing developmental status and potential benefits for hospital use. As a tool under continuous improvement, Savana Manager shows promise in improving hospital practices and patient care by enhancing clinical decisions. Building on this study, another promising future research involves leveraging our NLP algorithm to identify and classify patients with SpA from the broader hospital patient pool using extracted EHR features. Future endeavours will focus on expanding this research through multicentre studies and the integration of additional data elements, broadening the horizons of NLP applications in rheumatology.

In conclusion, this study adds preliminary evidence on the ability of Savana Manager, a tool designed for

the hospital setting that integrates NLP techniques, to analyse the complex nature of SpA. Our results emphasise the high prevalence of comorbidities and the use of treatments in SpA, yielding evidence on the ability of NLP to provide information from patients in real life. Finally, this methodology can be extrapolated to the rest of the medical pathologies. The ability of NLP to extract, organise and analyse vast amounts of unstructured EHR data may enable clinicians to have further insight from their daily practice. Furthermore, the integration of such tools into clinical support systems could help diagnostic processes, optimise treatment plans and ultimately improve patient outcomes by leveraging real-world data.

X Loreto Carmona @carmona_loreto

Acknowledgements Further investigators from the SAVANA Research Group are as follows: Iago Romero, Sebastian Menke, David Casadevall, Natalia Polo, and Guillermo Argüello.

Collaborators Savana Research group: Iago Romero, Sebastian Menke, David Casadevall, Natalia Polo and Guillermo Argüello.

Contributors DB: writing – original draft, data curation, investigation. JM-C and MT: conceptualisation, supervision, methodology, writing – review and editing. MBN: data curation, investigation, writing – review and editing. JAM: data curation, investigation, writing – review and editing. EdM: conceptualisation, supervision, methodology, writing – review and editing. All authors have contributed to this work and approved the final version. DB acts as the guarantor of this work. The manuscript used Savana Manager, an AI tool, to collect the data. During the preparation of this work, the author(s) used ChatGPT in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests DB: Speakers bureau: AbbVie, BMS, Galapagos, Janssen, Lilly, MSD. Research grants: Novartis. Consultancy: Sandoz, UCB. Part-time work in Savana Research. MBN worked at MedSavana during the development of the study. JMC and MT work at Savana Research SL. JAM and NS work at MedSavana. VNC: Consultancy/Speaker/Research grants from: Abbvie, BMS, Fresenius Kabi, Janssen, Lilly, MSD, Novartis, Pfizer, Roche, UCB. Member of ASAS Executive Committee. EdM: Research funding/consulting and conferences fees from: Abbvie, Novartis, Roche, Pfizer, Janssen, Lilly, MSD, BMS, UCB, Grunental and Sanofi.

Patient consent for publication Not applicable.

Ethics approval The project was approved by the Ethics Committee of Hospital La Paz, under the PI-5619.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The datasets generated and analysed during the current study are not publicly available and may be shared upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Diego Benavent <http://orcid.org/0000-0001-9119-5330>

María Benavent-Núñez <http://orcid.org/0000-0002-9248-748X>

Judith Marin-Corral <http://orcid.org/0000-0003-1320-4427>

Javier Arias-Manjón <http://orcid.org/0009-0008-3852-4092>
 Victoria Navarro-Compán <http://orcid.org/0000-0002-4527-852X>
 Miren Taberna <http://orcid.org/0000-0002-2446-186X>
 Diana Peiteado <http://orcid.org/0000-0002-6953-409X>
 Loreto Carmona <http://orcid.org/0000-0002-4401-2551>
 Eugenio de Miguel <http://orcid.org/0000-0001-5146-1964>

REFERENCES

- 1 Navarro-Compán V, Sepriano A, El-Zorkany B, et al. Axial spondyloarthritis. *Ann Rheum Dis* 2021;80:1511–21.
- 2 Sepriano A, Rubio R, Ramiro S, et al. Performance of the ASAS classification criteria for axial and peripheral spondyloarthritis: a systematic literature review and meta-analysis. *Ann Rheum Dis* 2017;76:886–90.
- 3 Moll JMH, Wright V. Psoriatic arthritis. *Semin Arthritis Rheum* 1973;3:55–78.
- 4 Classification of axial Spondyloarthritis inception cohort. ClinicalTrials.gov. Available: <https://clinicaltrials.gov/study/NCT03993847> [Accessed 5 Jan 2024].
- 5 Poddubnyy D, Baraliakos X, Van den Bosch F, et al. Axial involvement in psoriatic arthritis cohort (AXIS): the protocol of a joint project of the assessment of spondyloarthritis international society (ASAS) and the group for research and assessment of psoriasis and psoriatic arthritis (GRAPPA). *Ther Adv Musculoskelet Dis* 2021;13.
- 6 Bajwa J, Munir U, Nori A, et al. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J* 2021;8:e188–94.
- 7 Knevel R, Liao KP. From real-world electronic health record data to real-world results using artificial intelligence. *Ann Rheum Dis* 2023;82:306–11.
- 8 Venerito V, Bilgin E, Iannone F, et al. AI am a rheumatologist: a practical primer to large language models for rheumatologists. *Rheumatology* 2023;62:3256–60.
- 9 Zhao SS, Hong C, Cai T, et al. Incorporating natural language processing to improve classification of axial spondyloarthritis using electronic health records. *Rheumatology* 2020;59:1059–65.
- 10 Humbert-Droz M, Izadi Z, Schmajuk G, et al. Development of a natural language processing system for extracting rheumatoid arthritis outcomes from clinical notes using the National rheumatology informatics system for effectiveness registry. *Arthritis Care & Research* 2023;75:608–15.
- 11 Canales L, Menke S, Marchesseau S, et al. Assessing the performance of clinical natural language processing systems: development of an evaluation methodology. *JMIR Med Inform* 2021;9:e20492.
- 12 Graziani D, Soriano JB, Del Rio-Bermudez C, et al. Characteristics and prognosis of COVID-19 in patients with COPD. *J Clin Med* 2020;9:3259.
- 13 Izquierdo JL, Almonacid C, González Y, et al. The impact of COVID-19 on patients with asthma. *Eur Respir J* 2021;57:2003142.
- 14 Larrainzar-Garjón R, Fernández-Tormos E, Collado-Escudero CA, et al. Predictive model for a second hip fracture occurrence using natural language processing and machine learning on electronic health records. *Sci Rep* 2024;14:532:532.
- 15 González-Juanatey C, Anguita-Sánchez M, Barrios V, et al. Major adverse cardiovascular events in coronary type 2 diabetic patients: identification of associated factors using electronic health records and natural language processing. *J Clin Med* 2022;11:6004.
- 16 Benson T. *Principles of health interoperability HL7 and SNOMED: Second Edition*. 2012:1–316.
- 17 UMLS metathesaurus - Snomedct_Us (SNOMED CT, US edition) - synopsis. Available: https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/SNOMEDCT_US/index.html [Accessed 14 Dec 2023].
- 18 WMA declaration of helsinki – ethical principles for medical research involving human subjects – WMA – the World Medical Association. Available: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/> [Accessed 15 Feb 2024].
- 19 Seoane-Mato D, Sánchez-Piedra C, Silva-Fernández L, et al. Prevalencia de enfermedades reumáticas en población adulta en España (Estudio EPISER 2016). *Objetivos Y metodología. Reumatología Clínica* 2019;15:90–6.
- 20 Dean LE, Jones GT, MacDonald AG, et al. Global prevalence of ankylosing spondylitis. *Rheumatology* 2014;53:650–7.
- 21 Passia E, Vis M, Coates LC, et al. Sex-specific differences and how to handle them in early psoriatic arthritis. *Arthritis Res Ther* 2022;24.
- 22 Chimenti M-S, Alten R, D'Agostino M-A, et al. Sex-associated and gender-associated differences in the diagnosis and management of axial spondyloarthritis: addressing the unmet needs of female patients. *RMD Open* 2021;7:e001681.
- 23 Benavent D, Capelusnik D, Ramiro S, et al. Does gender influence outcome measures similarly in patients with Spondyloarthritis? Results from the ASAS-perSpA study. *RMD Open* 2022;8:e002514.
- 24 Benavent D, Navarro-Compán V. Understanding the paradigm of non-radiographic axial spondyloarthritis. *Clin Rheumatol* 2021;40:501–12.
- 25 López-Medina C, Molto A, Sieper J, et al. Prevalence and distribution of peripheral musculoskeletal manifestations in spondyloarthritis including psoriatic arthritis: results of the worldwide, cross-sectional ASAS-perspa study. *RMD Open* 2021;7:e001450.
- 26 Bautista-Molano W, Landewé R, Burgos-Vargas R, et al. Prevalence of comorbidities and risk factors for comorbidities in patients with spondyloarthritis in Latin America: a comparative study with the general population and data from the ASAS-COMOSPA study. *J Rheumatol* 2018;45:206–12.
- 27 Puche-Larrubia MÁ, Ladehesa-Pineda L, Gómez-García I, et al. Impact of the number of comorbidities on the outcome measures and on the retention rate of the first anti-TNF in patients with ankylosing spondylitis. *Semin Arthritis Rheum* 2022;52:151938.
- 28 Kroon FPB, van der Burg LRA, Ramiro S, et al. Nonsteroidal antiinflammatory drugs for axial spondyloarthritis: a cochrane review. *J Rheumatol* 2016;43:607–17.
- 29 Wang R, Bathon JM, Ward MM. Nonsteroidal antiinflammatory drugs as potential disease-modifying medications in axial spondyloarthritis. *Arthritis Rheumatol* 2020;72:518–28.