Supplementary Table 1. SNOMED terms for the SpA variable

Concept id	SNOMED term- Spanish search
Spondyloarthritis	
201575009	artropatía reactiva de primera articulación metatarsofalángica
201567003	artropatía reactiva de la cadera
422565003	poliartritis posinfecciosa
201563004	artropatía reactiva de la muñeca
201481004	artropatía reactiva de transmisión sexual, de múltiples localizaciones
239804002	artritis psoriásica juvenil sin psoriasis
2194	Espondiloartritis periférica
201570004	artropatía reactiva de la articulación tibioperonea
239810002	espondilitis anquilosante con compromiso orgánico/sistémico
239785008	enfermedad de Reiter posdisentérica
56528004	artropatía posdisentérica
2195	Espondiloartritis axial no radiográfica
201805000	artritis juvenil en la enfermedad de Crohn
81077008	artritis reumática aguda
201807008	artritis juvenil en la colitis ulcerosa
201566007	artropatía reactiva de articulación interfalángica distal de dedo de la mano
239783001	artritis posinfecciosa
239808004	síndrome de Reiter juvenil
201474006	artropatía reactiva de región de hombro adquirida por vía sexual
201477004	artropatía reactiva de transmisión sexual, de la mano
240428000	artritis posmeningocócica
201497009	artropatía reactiva posdisentérica de región de hombro
201738001	artropatía posinfecciosa en la sífilis
239806000	espondiloartropatía juvenil
201500003	artropatía reactiva posdisentérica, de la mano
2154	Espondilitis indiferenciada periférica
239811003	espondilitis anquilosante con compromiso multisistémico
201559006	artropatía reactiva de la articulación esternoclavicular
423515003	artritis posinfeciosa de articulación del pie
201558003	artropatía reactiva de hombro
201561002	artropatía reactiva del codo
123949001	artritis reactiva posestreptocócica
19514005	artritis mutilante
429422002	artritis reumática de articulación temporomandibular
201501004	artropatía reactiva posdisentérica, de región pélvica y muslo
2148	Artritis asociada a EIIC
201504007	artropatía reactiva posdisentérica, de múltiples localizaciones
201573002	artropatía reactiva de la articulación astragalonavicular
410482007	iritis en la artritis psoriásica
67224007	enfermedad de Reiter

201480003	artropatía reactiva de transmisión sexual, del tobillo y el pie
239813000	dactilitis psoriásica
239812005	artritis psoriásica con compromiso de la articulación interfalángica distal
201569000	artropatía reactiva de la rodilla
201576005	artropatía reactiva de la quinta articulación metatarsofalángica
239805001	espondilitis anquilosante juvenil
239787000	artritis reactiva postuberculosa
201562009	artropatía reactiva de la articulación radiocubital distal
200956002	artritis psoriásica con compromiso vertebral
415141001	infección genitourinaria consecutiva a enfermedad de Reiter
9631008	espondilitis anquilosante
239802003	artritis psoriásica juvenil
423310007	poliartritis posinfecciosa de articulación de la mano
239803008	artritis psoriásica juvenil con psoriasis
201564005	artropatía reactiva de la articulación metacarpofalángica
239809007	artritis juvenil en la enfermedad intestinal inflamatoria
2149	Espondiloartritis axial y periférica
129564003	espondiloartritis anquilosante y lesiones oculares
14175009	fiebre reumática articular
2153	Espondilitis indiferenciada axial
162930007	al examen: deformidad torácica de la espondilitis anquilosante
201572007	artropatía reactiva de la articulación subastragalina
201571000	artropatía reactiva del tobillo
2150	Espondiloartropatía HLA B 27 positivo
201503001	artropatía reactiva posdisentérica, del tobillo y el pie
267883008	artritis reactiva postinfección genitourinaria
58769002	artritis reumática subaguda
201565006	artropatía reactiva de articulación interfalángica proximal de dedo de la mano
201577001	artropatía reactiva de articulación interfalángica de dedo del pie
201568008	artropatía reactiva de la articulación sacroilíaca
2151	Espondiloartropatía HLA B 27 negativo
201560001	artropatía reactiva de la articulación acromioclavicular
2193	Espondiloartritis axial
9350004	artritis enteropática
201478009	artropatía reactiva de transmisión sexual, de la región pélvica y del muslo
2152	Espondiloartropatía indiferenciada
33339001	psoriasis con artropatía

Supplementary Table 2. Analysis stratified by time-windows

Variable	Period 1*	Period 2**	Z-statistic
Total number of patients in the hospital	422,100	473,906	
Total number of SpA patients	2,725	2,988	0.89
Demographics and toxic habits			
Sex (male)	1,281 (47.01%)	1,309 (43.81%)	2.43
Type of involvement			
Peripheral Spondyloarthritis*	881 (32.33%)	941 (31.49%)	0.68
Axial Spondyloarthritis	653 (23.96%)	694 (23.23%)	0.66
Axial and Peripheral Spondyloarthritis	56 (2.06%)	44 (1.47%)	1.68
Extra-musculoskeletal manifestations			
Uveitis	598 (21.94%)	786 (26.31%)	-3.84
Inflammatory bowel disease	281 (10.31%)	326 (10.91%)	-0.73
Psoriasis	818 (30.02%)	958 (32.06%)	-1.67
Peripheral manifestations			
Enthesitis	548 (20.11%)	663 (22.19%)	-1.92
Dactylitis	207 (7.60%)	227 (7.60%)	-0.001
Cardiovascular risk factors			
Active smoker	353 (12.95%)	390 (13.05%)	-0.11
High blood pressure	595 (21.83%)	616 (22.62%)	1.13
Diabetes Mellitus	308 (11.30%)	309 (11.34%)	1.17
Dyslipidemia	471 (17.28%)	523 (19.19%)	-0.22
Treatment			
Methotrexate	786 (28.84%)	861 (28.82%)	0.02
Sulfasalazine	403 (14.79%)	438 (14.66%)	0.14
Adalimumab	295 (10.83%)	359 (12.02%)	-1.41
Infliximab	179 (6.57%)	190 (6.36%)	0.32
Etanercept	204 (7.49%)	207 (6.93%)	0.82

*From January 1st 2020 to April 30th 2021; ** From May 1st 2021 to 21st August 2022

Supplementary Text 1

Clinical NLP processing

Savana Manager contains a subset of functionalities of EHRead® technology mainly consisting of a layer of named-entity recognition including named-entity disambiguation plus baseline ML models covering for example negation and temporality detection.

The SpA related variables were detected in the free text using a named-entity recognition approach. This initial step identifies clinical entities within the text, distinguishing them from non-clinical content. Once identified, entities are classified into distinct categories. As additional layers, negation, temporality and section detection were applied. Our general models are trained using standard terminologies such as SNOMED-CT or ATC and meticulously improved through clinician annotations to capture the most relevant expressions for each variable that is not available in the standard version of the terminologies. Therefore, common typos and medical abbreviations are captured, significantly improving their capability to accurately interpret the diverse and often inconsistent terminology found within EHR data.

Negation and speculation recognition model combines rule-based methods for detecting common negation triggers with neural networks capable of handling complex negation structures (1), which has been trained on real Spanish EHRs and evaluated against a rich set of reference standards. This model classifies each clinical entity as being affirmative or non-affirmative based on its lexical and semantic context.

The temporality detection is carried out by a NLP module that consists of various layers that work in combination to assign dates to clinical entities. The first layer is a named-entity detection engine responsible for the detection of any mentioning of dates in the free text of EHRs. Subsequently, a relationship model based on a Bi-LSTM (2) decides if a detected date is related to a detected clinical entity. In addition, a normalization layer takes care of converting different date formats as written in the EHRs' free text into a normalized representation.

As for section detection, the model categorizes sentences within the EHR text into predefined sections (e.g., patient history, current conditions), allowing for a better interpretation of the data.

The last step of the NLP processing, the post-processing step, carries out several quality control operations and combines the output from the different NLP modules into a final database.

1- Tay Y, Dehghani M, Gupta J, Bahri D, Aribandi V, Qin Z, et al. Are Pre-trained Convolutions Better than Pre-trained Transformers? [Internet]. arXiv; 2022 [cited 2022 Aug 16]. Available from: http://arxiv.org/abs/2105.03322

2- Schuster, M. and Paliwal, K.K. (1997) Bidirectional Recurrent Neural 1- Networks. IEEE Transactions on Signal Processing, 45, 2673-2681.2-https://doi.org/10.1109/78.650093

Supplementary text 2

Before the data could leave the Hospital Universitario La Paz to Savana's servers via a secure route, an important procedure needed to be executed beforehand during which personal health information was pseudonymized (protected by a pseudonymization step). The pseudonymization process consisted of utilizing a hashing tool for pseudonymization of identifiers to transform identifiers into a set of characters, including patient and episode identifiers, and it was conducted at the Hospital Universitario La Paz. Medsavana supplied the hospital with the necessary tools and support for pseudonymizing the information contained within the EHRs, with Medsavana taking on the role of data processing for this purpose at the hospital.

Given that the hospital performed the pseudonymization of this data before sending it to MedSavana, any supplementary information that could potentially de-anonymize the data and thus reveal patient identities was always kept separated, with MedSavana having no access to it. Following the conclusion of this initial pseudonymization stage, the database was then forwarded to MedSavana's systems.

Moreover, the process focused on the extraction and analysis of clinical variables, and free text from medical records is excluded from the platform, minimizing privacy and confidentiality risks.