PROLEPSIS

D3.1 / PsA inflammation drivers research

Editor	Contractual delivery date	Actual delivery date
Batoul Hojeij (CICERO)	May 2025	July 2025
Deliverable type	**Dissemination level**	**Version - date**
R - Document, report	PU – Public	1.0 - 31/07/2025

Funded by the
European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the European Health and Digital Executive Agency can be held responsible for them.

Deliverable ID

Project acronym	iPROLEPSIS
Project full title	Psoriatic arthritis inflammation explained through multi-source data analysis guiding a novel personalised digital care ecosystem
Grant Agreement ID	101095697

Deliverable number	D3.1
Deliverable title	PsA inflammation drivers research
Work package	WP3 - Research on PsA inflammation drivers and monitoring
Deliverable type	R - Document, report
Dissemination level	PU – Public
Version - date	1.0 - 31/07/2025

Contractual delivery date	May 2025
Actual delivery date	July 2025

Lead partner	CICERO
Editor	Batoul Hojeij (CICERO)
Contributors	Batoul Hojeij (CICERO); Ilja Tchetverikov (CICERO); Eleni Vasileiou (AUTH); Giorgos Apostolidis (AUTH); Vasilis Charisis (AUTH); Nikos Melanitis (AIN); Jolanda Luime (CICERO); Cátia Gonçalves (PSR)
Reviewed by	Laura Coates (UOXF) Francesca Levi-Schaffer (HUJI)
Approved by	Leontios Hadjileontiadis (AUTH, Project Coordinator)

Keywords	Disease activity; flare; inflammation; prognosis; psoriatic arthritis

Document history

Version	Date	Contributors	Action / status
0.1	12/07/2024	CICERO	Document structure (table of contents) ready
0.2	13/06/2025	CICERO/EMC/AUTH/AIN/PSR	Ready for internal review
0.2	20/06/2025	UOXF	Reviewed by Laura Coates (UOXF)
0.2	22/06/2025	HUJI	Reviewed by Francesca Levi-Schaffer (HUJI)
0.3	11/07/2025	CICERO/AIN	Document revised
0.4	29/07/2025	AUTH	Minor issues resolved and formatting issues fixed; Approved by the Project Coordinator
1.0	31/07/2025	AUTH	Submitted to the EC by the Project Coordinator

Contents

List of figures

List of tables

List of abbreviations

ANOVA	Analysis of Variance
API	Application Programming Interface
BSA	Body Surface Area
BMI	Body Mass Index
CDF	Cumulative Distribution Functions
CRP	C-Reactive Protein
CVD	Cardiovascular Disease
DAPSA	Disease Activity in Psoriatic Arthritis
DEPAR	Dutch South West Psoriatic Arthritis cohort
DMARDs	Disease-Modifying Anti-Rheumatic Drugs
EU	European
EULAR	European Alliance of Associations for Rheumatology
GC	Glucocorticoids
GBTM	Group-Based Trajectory Models
GRACE	GRAPPA Composite Exercise
HADS	Hospital Anxiety and Depression Scale
HAQ	Health Assessment Questionnaire
HDI	Highest Density Interval
HLA	Human Leukocyte Antigen
HR	Hazard Ratio
IBD	Inflammatory Bowel Disease
ISCED	International Standard Classification of Education
IQR	Interquartile Range
kNN	k-Nearest Neighbors
LDA	Low Disease Activity
LEI	Leeds Enthesitis Index
MDA	Minimal Disease Activity
MREC	Medical Research Ethics Committee
MTX	Methotrexate
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
NUTS	No-U-Turn Sampler
OOF	Out-of-Fold
OR	Odds Ratio
PASDAS	Psoriatic Arthritis Disease Activity Score
PASS	Patient Acceptable Symptom State
PASI	Psoriasis Area and Severity Index
PDPID	PsA digital phenotyping and inflammation drivers study
PRO	Patient-Reported Outcome
PSS	Perceived Stress Scale
PsA	Psoriatic Arthritis
PsAID	Psoriatic Arthritis Impact of Disease
PsO	Psoriasis
SD	Standard Deviation
SF-36	Short Form 36 Health Survey

SF36 PCS	Short Form 36 Physical Component Score
SHAP	SHapley Additive exPlanations
SJC	Swollen Joint Count
TJC	Tender Joint Count
VAS	Visual Analogue Scale
WP	Work Package

Executive summary

Psoriatic arthritis (PsA) is a chronic inflammatory disease associated with psoriasis, and involves interactions of genetic, immunologic, and environmental factors. Patients with PsA can differ in their disease manifestation and treatment response. PsA prognosis also impacts the patient quality of life. Besides, patients with PsA can experience periods of increased disease activity, known as flares, which can negatively impact their quality of life. Currently, literature on the triggers of these flares is very limited.

Deliverable D3.1 is the initial report of Task T3.1 that aims to identify key drivers of inflammation in PsA. This objective is achieved through the analysis of retrospective datasets to which the consortium has access, as well as datasets collected during a prospective study, namely PDPID. This deliverable presents the outcomes of the exploratory analysis of two retrospective datasets, i.e., DEPAR and Reuma.pt, and a subset of the PDPID prospective data.

Using retrospective data analysis of DEPAR and Reuma.pt cohorts, we addressed the PsA population characteristics in both cohorts, as well as sex differences in PsA-related measures and timing of diagnosis and disease prognosis. Our analysis showed that there are differences between the DEPAR and Reuma.pt cohort patients in variables such as disease activity, duration and phenotype, and treatment. When comparing males and females, differences were observed across multiple variables such as disease activity, duration and phenotype, joint count, pain levels, treatment duration and retention. Concerning timing of diagnosis and disease prognosis, we showed that early PsA diagnosis is associated with lower disease activity and better quality of life measures. On the other hand, our results suggest that diagnosis timing alone does not determine treatment trajectory of the patients.

The PDPID study primarily aims are to digitally monitor flare and identify its triggers. We performed preliminary analysis of a cross-section of patient baseline clinical data from Netherlands, UK, and Greece. Preliminary analysis results reveal that 17% of patients experienced flare according to the physician while 10% of patients reported experiencing a flare. There were differences between countries in many variables related to disease activity, clinical assessment, and patient-reported outcomes. The flare measures (flare, remission, low disease activity) were correlated. Flare was associated with higher pain, fatigue, sleep disturbance, and stiffness levels. Of interest, higher perceived stress was associated with higher disease activity. Moreover, higher nitrogen oxide exposure was associated with higher disease activity, while higher ozone exposure was associated with lower disease activity. As a future step, the analysis will be expanded to the four countries (Netherlands, UK, Greece, and Portugal) with more participants and follow-up data and including additional clinical data (clinical measures, biological data, environmental data), to achieve the study objectives and identify triggers of inflammation.

The outcomes of the aforementioned analyses, along with PsA-specific digital biomarkers (to be presented in “D3.2 – First report on digital phenotyping for PsA”), will be used to train the flare detection and prediction AI models (to be presented in “D3.4-First report on data modelling”).

1 Introduction

Psoriatic arthritis (PsA) is a chronic inflammatory disease associated with psoriasis, with a prevalence of 0.1% in the general population (Lembke et al., 2024). PsA involves interactions between genetic, immunologic, and environmental factors (Veale & Fearon, 2018). Patient with PsA can experience exacerbations in their disease activity, known as flares, which can impact their quality of life (Orbai et al., 2023). Although the pathophysiology of PsA is understood, drivers of inflammation leading to these flares remain to be elucidated (Veale & Fearon, 2018).

PsA prognosis can impact the patient physical, psychological, and social well-being (Husni et al., 2017; van Hal et al., 2022; Walsh et al., 2023). Several factors can influence PsA prognosis, including disease activity, the severity and extent of joint and skin involvement, the presence of comorbidities (e.g., cardiovascular diseases), and the response to treatment (Cañete et al., 2020; Mekhail et al., 2020; Tillett et al., 2020; van Hal et al., 2022; Walsh et al., 2023). For example, patients with active disease are at higher risk for joint damage, functional impairment and reduced quality of life (Gottlieb et al., 2023; Snoeck Henkemans et al., 2022).

Genetic factors play a role in the prognosis of PsA. For example, certain genetic markers and polymorphisms are associated with specific disease phenotypes (e.g., HLA-B*27, rs1800795) and can contribute to the variability in disease activity (e.g., rs33980500, rs16944) (Cubino et al., 2016; De Benedittis et al., 2022; Queiro et al., 2002). The gut microbiome is closely connected to the immune system and inflammation, which are core pathways affecting PsA disease activity, suggesting that gut dysbiosis can trigger inflammation in patients with PsA (Maciel-Fiuza et al., 2023; Veale & Fearon, 2018). However, it is currently unknown how genetics and gut microbiome contribute to flare in PsA. Furthermore, environmental factors such as unhealthy lifestyle behaviours (e.g., stress and sleep disorders), air pollution, fatigue, and pain can trigger inflammation and immune response, suggesting that they can be implicated in PsA flare (Baral et al., 2019; Besedovsky et al., 2019; Dhabhar, 2014; Pedraz-Petrozzi et al., 2020; Tripathy et al., 2021).

Patients with PsA can differ in their disease manifestation and treatment response (Gratacós et al., 2021; McArdle et al., 2018). Sex differences are one factor that can help explain these differences (Tarannum et al., 2022). Men and women may experience PsA differently, including differences in disease activity, response to treatment, and overall disease burden (Duruöz et al., 2021; Tarannum et al., 2022). However, the reasons behind these differences are not yet fully understood (Tarannum et al., 2022).

The DEPAR, MONITOR, and Reuma.pt are existing cohorts of patients with PsA in the Netherlands, UK and Portugal, respectively. The DEPAR cohort includes patients newly diagnosed with PsA, whereas Reuma.pt is a national registry that includes patients living with rheumatic and musculoskeletal diseases, including patients with PsA. These cohorts aim to enhance the understanding of PsA and improve patient care by collecting information on their medical history, clinical features, disease activity, medication and patient-reported outcomes (PROs).

The PDPID study is a prospective cohort that includes patients with PsA from Netherlands, UK, Portugal and Greece. The primary aims are to digitally monitor and identify the triggers of flare in patients with PsA. Besides digital data collection, the PDPID collects data on flare and other disease activity measures, clinical assessments, patient-reported outcomes, medical history, DNA, hair cortisol, gut microbiome and environmental data.

1.1 Document scope

The deliverable is the initial report for Task T3.1, part of work package 3 (WP3). This task deals with identifying the drivers of inflammation in PsA. In this deliverable, we report on the retrospective data analysis from the clinical cohorts DEPAR and Reuma.pt, to address three topics: characterization of the PsA population, sex differences and disease prognosis. In addition, we provide preliminary results for the analysis of clinical data collected in the PDPID study and the data analysis plan. The results of the analysis will inform the development of models in Task T3.5.

1.2 Document structure

This document is a comprehensive report on the research into inflammation drivers in PsA. The document consists of four sections: (1) introduction, (2) retrospective data analysis from the DEPAR and Reuma.pt cohorts, (3) prospective analysis of the PDPID cohort data, and (4) conclusion. Section 1 includes the document introduction, scope and structure. In Section 2, the objectives, methods, and results of the retrospective data analysis for three research questions are provided. Section 3 provides information on the data analysis plan and preliminary results for the clinical collected in the PDPID study. Finally, Section 4 concludes by summarizing the main points of the document and outlining future steps.

2 Retrospective data analysis: DEPAR and Reuma.pt cohorts

2.1 Objectives

Using existing data collected from the DEPAR and Reuma.pt cohorts we aim to:

(1) Describe and compare the baseline characteristics of patients with PsA across the two participating cohorts (DEPAR and Reuma.pt) to identify any statistically significant differences between them.

(2) Investigate the sex differences in PROs, disease activity states and phenotypes in patients with PsA. The analysis explores baseline characteristics of male and female patients, tracking changes over time in disease activity and PROs, and assesses how disease progression trajectories vary by sex.

(3) Investigate associations between late diagnosis of PsA and disease prognosis, in terms of disease activity and phenotype, disability, quality of life and treatment persistence.

2.2 Methods

2.2.1 Patients and settings

This is a retrospective analysis of data from patients newly diagnosed with PsA included from DEPAR (Netherlands) and Reuma.pt (Portugal) cohorts. In DEPAR, patients were included between 2013-2023 and in Reuma.pt patients were included between 2008-2024. Only newly diagnosed patients were included from Reuma.pt, i.e., those with a baseline visit within one year of being diagnosed with PsA. Each cohort was approved from its local medical ethics review committees, and all patients provided informed consent. Of note, the MONITOR dataset (UK) was not included in the analysis due to delays in delivery of the dataset.

The DEPAR cohort is a multi-center prospective study that includes sites in the southwestern region of the Netherlands (Kasiem et al., 2021). The cohort includes patients newly diagnosed

with PsA according to their treating physician, are at least 18 years old, have not started treatment with disease-modifying antirheumatic drugs (DMARDs) for their joint symptoms at the time of inclusion, and have sufficient knowledge of the Dutch language.

Reuma.pt is the Portuguese Registry of Rheumatic Diseases, an observational, prospective, long-term national registry promoted by the Portuguese Society of Rheumatology (SPR). It collects health data from individuals living with rheumatic and musculoskeletal diseases across 76 centres in Portugal (mainland, Azores and Madeira), primarily public hospitals with rheumatology departments. Active since 2008, Reuma.pt supports prospective data collection through disease-specific protocols for 14 musculoskeletal and rheumatic diseases, including rheumatoid arthritis, psoriatic arthritis, spondylarthritis, juvenile idiopathic arthritis, systemic lupus erythematosus, autoinflammatory syndromes, osteoporosis/osteoporotic fractures, early arthritis, vasculitis, osteoarthritis, scleroderma, Sjögren's syndrome, myositis, and mixed connective tissue diseases. In addition, two generic protocols are available for other rheumatic diseases affecting both children and adults. Participation in Reuma.pt requires written informed consent.

2.2.2 Demographics and medical history

Patients demographics included age, sex, education, employment status were assessed at baseline. Disease duration, HLA-B27 status, body mass index (BMI), smoking behaviour, and disease manifestations and comorbidities (uveitis, irritable bowel disease (IBD), cardiovascular diseases (CVD)) were assessed as part of patient medical history (**Table 1**).

2.2.3 Clinical assessment

Clinical examination included assessment for disease phenotype, and 66 swollen and 68 tender joint count (SJC 66 and TJC 68, respectively), body surface area (BSA)/PASI, Leeds enthesitis index (LEI) for enthesitis (Healy & Helliwell, 2008) and dactylitis counting score (**Table 1**). We assign patients to early or late diagnosis group based on the time interval τ_d between onset of symptoms and PsA diagnosis. The cutoff values used in the analysis are: Early diagnosis if $\tau_d \leq 1year$ and Late diagnosis if $\tau_d > 1year$. We also try splitting into four diagnosis groups, i.e. Very Early: $\tau_d \leq 3months$, Early: $3months < \tau_d \leq 1year$, Intermediate: $1year < \tau_d \leq 2years$ and Late: $\tau_d > 2years$, diagnosis groups.

Table 1 Data assessed at each time-point from DEPAR and/or Reuma.pt

Variable	Parameter	T0	T3	T6	T9	T12	T18	T24	T36
Age	Demographics	X							
Sex	Demographics	X							
Education	Demographics	X							
Employment status	Demographics	X				X	X	X	X
Disease duration	Medical history	X							
HLA-B27	Medical history	X							
BMI	Medical history	X				X	X	X	X
Smoking behaviour	Medical history	X				X	X	X	X
Comorbidities (IBD, CVD)	Medical history	X				X	X	X	X
Disease phenotype	Clinical assessment	X							
SJC/TJC 66/68	Clinical assessment	X	X	X	X	X	X	X	X

Variable	Parameter	T0	T3	T6	T9	T12	T18	T24	T36
LEI	Clinical assessment	X	X	X	X	X	X	X	X
BSA/PASI	Clinical assessment	X	X	X	X	X	X	X	X
Dactylitis count	Clinical assessment	X	X	X	X	X	X	X	X
Symptom onset (patient)	Clinical assessment	X							
PASDAS	Disease activity	X	X	X	X	X	X	X	X
DAPSA	Disease activity	X	X	X	X	X	X	X	X
MDA	Disease activity	X	X	X	X	X	X	X	X
CRP	Disease activity	X	X	X	X	X	X	X	X
Medication	Treatment	X	X	X	X	X	X	X	X
HAQ	PROS	X	X	X	X	X	X	X	X
SF-36	PROS	X	X	X	X	X	X	X	X
VAS pain	PROS	X	X	X	X	X	X	X	X
VAS global	PROS	X	X	X	X	X	X	X	X
HADS anxiety/depression	PROS	X	X	X	X	X	X	X	X
GRACE	PROS	X	X	X	X	X	X	X	X

2.2.4 Disease activity and impact of disease assessment

Disease activity was assessed at baseline and during the follow-up assessments. Disease activity measures included composite disease activity scores: Disease Activity index for PsA (DAPSA) and Psoriatic Arthritis Disease Activity Score (PASDAS), and whether the patient was in minimal disease activity (MDA) state (**Table 2**) (Sandoval & Fernández-Ávila, 2023). CRP was also measured and included as part of disease activity (**Table 1**).

Table 2. Disease activity scores calculation

Disease activity score	Equation
MDA	5 of 7 of the following criteria: • TJC ≤1 • SJC ≤1 • PASI ≤1 or BSA ≤3 • Patient pain VAS ≤15 • Patient global activity VAS ≤20 • HAQ ≤0.5 • Tender entheseal points ≤1
PASDAS	PASDAS = (((0.18 × √physician global VAS) + (0.159 × √patient global VAS) – (0.253 × √SF36 – PCS) + (0.101 × LN (SJC + 1)) + (0.048 × LN (TJC + 1)) + (0.23 × LN (Leeds Enthesitis Index + 1)) + (0.377 × LN (tender dactylitis count + 1)) + (0.102 × LN (CRP* + 1)) +2) × 1.5
DAPSA	SJC66 + TJC68 + PtGA (patient global assessment VAS) + patient pain VAS + CRP**

* CRP in mg/L; **CRP in mg/dl

2.2.5 Treatment outcomes

Medications were grouped into four categories: (i) cs-DMARD (ii) b-DMARD (iii) ts-DMARD (iv) Methotrexate.

Outcomes rated to treatment included time to start medication, time to reach stable disease activity following medication and drug survival. Treatment persistence was defined as sticking to the same medication (i.e., have not changed therapy or failed multiple biologics).

2.2.6 Patient-reported outcomes

PROs were assessed using questionnaires at baseline and during the follow-up assessments. These included the Health Assessment Questionnaire (HAQ) to measure physical function (Fries et al., 1980), Visual Analog Scales (VAS) pain and VAS global, and HADS for anxiety and depression (

Table 1).

2.2.7 Data analysis

The variables analysed for each objective are provided in **Table 3**.

Table 3 Variables analysed for each objective (DEPAR, and Reuma.pt)

Variable	Objective 1	Objective 2	Objective 3
Age	X	X	X
Sex	X	X	X
Education	X	X	
Employment status	X	X	
Disease duration	X	X	
HLA-B27	X	X	
BMI	X	X	X
Smoking behaviour	X	X	X
Comorbidities (IBD, CVD)	X	X	
Disease phenotype	X	X	X
SJC/TJC 66/68	X	X	
LEI	X	X	
BSA/PASI	X	X	
Dactylitis count	X	X	
Disease duration	X	X	
Diagnosis			X
PASDAS	X	X	X
DAPSA	X	X	X
MDA	X	X	X
CRP	X	X	
HAQ	X	X	
SF-36	X	X	X
VAS pain	X	X	
VAS global	X	X	

Variable	Objective 1	Objective 2	Objective 3
HADS anxiety/depression	X	X	
GRACE	X	X	X
Time to start medication		X	
Time to reach stable disease activity following medication		X	
Drug survival		X	
Treatment persistence			X

2.2.7.1 Objective 1

Baseline data for continuous variables, such as age, disease duration, and PROs, were summarized using the mean and standard deviation (SD) to provide an overview of central tendencies and variability within each cohort. For categorical variables, such as sex, disease phenotypes, and treatment types, frequencies and percentages were used to illustrate the distribution of patients across categories.

To assess whether there were statistically significant differences between the cohorts at baseline, statistical comparisons were performed. Continuous variables were compared using either analysis of variance (ANOVA) or Kruskal-Wallis tests, depending on the data's distribution. Categorical variables were analysed using Chi-square tests or Fisher's exact tests when applicable.

2.2.7.2 Objective 2

Baseline data for the study were summarized using mean and SDs (or median and interquartile ranges (IQR)) for continuous variables, while categorical variables, including disease phenotypes, were described using frequency and percentages. Statistical comparisons of these baseline characteristics between different sexes were performed at two time points: baseline and at 36 months (T36). To assess changes over time in disease activity and PROs, we modeled longitudinal data using either linear mixed-effects models (for continuous outcomes like DAPSA) or generalized estimating equations (for binary outcomes like achieving Minimal Disease Activity (MDA)). Both approaches accounted for repeated measurements within individuals. Models included sex, time, age, cohort, and a sex-by-time interaction to assess differential change over time by sex.

Treatment patterns were analyzed descriptively by summarizing medication usage at each follow-up visit. Kaplan–Meier survival curves were used to estimate treatment retention rates for each medication class, stratified by sex. Log-rank tests were used to assess sex differences in retention over 36 months. Medication records were harmonized across cohorts and categorized into medication classes. Treatment intervals were aligned with each patient's observation window, and overlapping exposures were merged to ensure accurate cumulative duration.

2.2.7.3 Objective 3

In this section we give brief outlines of the methods used. More details are available in the **Appendix I.1.**

2.2.7.3.1 Statistical tests

To assess the impact of diagnostic timing on disease activity progression, we performed independent two-sample t-tests and non-parametric Mann-Whitney U tests, to compare the mean disease activity scores between early and late diagnosis groups. This analysis was

conducted at each scheduled visit (baseline [M0], 3 months [M3], 6 months [M6], and M12, M18, M24, M36 visits) across multiple outcome measures, specifically PASDAS, DAPSA, GRACE, and both overall and subdomain scores of the SF-36. The aim was to determine whether there were statistically significant differences in disease activity trajectories between the two groups over time. For each disease activity measure, we conducted two sample t-tests comparing group means at successive visits from baseline (First Visit) to 36 months (T36). The analysis employed summary statistics (mean, standard deviation, and sample size), ensuring robustness despite unequal group sizes using Welch's t-test. When normality (assessed via the Lilliefors test) could not be assumed, Mann-Whitney U test was used.

2.2.7.3.2 Bayesian Linear Regression

We employed a unified Bayesian linear regression framework to model the progression of disease activity over time in patients with psoriatic arthritis, stratified by diagnostic timing (*very early*, *early*, *intermediate*, *late*). Separate models were fit for different outcome measures, reflecting their underlying scale and distributional characteristics. In particular, we distinguished between continuous scores (e.g., PASDAS) and categorical-like outcomes (e.g., MDA).

Data and Predictors

All models used harmonized visit-level clinical data from standardized timepoints (e.g., T0, T3, ..., T36 months). The core predictors included:

- Time: Numeric variable (in months) representing follow-up duration from baseline, scaled for numerical stability.
- Diagnosis group: Categorical variable encoded as an integer.
- Interaction: Time-by-diagnosis group interaction to allow differential temporal dynamics across diagnosis strata.

These predictors were broadcast over the patient-time grid using array manipulation. For each outcome, a separate model was fit using a consistent structural template but adapted for the outcome's statistical properties.

In the continuous outcome model, "Change in disease score from baseline" (ΔPASDAS) was used as the outcome of the Bayesian Linear Regression Model. In the discrete outcome model, MDA was used as the outcome of the Bayesian Logistic Regression Model

We employed weakly informative priors throughout the model to provide regularization while retaining flexibility. Missing outcome values (Y) were handled natively by the probabilistic programming framework: any observations with missing values were automatically excluded from the likelihood without requiring imputation, preserving uncertainty

2.2.7.3.3 Treatment Escalation: Clustering, COX model and kNN Feature Engineering

EULAR guidelines

Treatment escalation in PsA was investigated based on the 2023 EULAR recommendations [i] for pharmacological management (see **Figure 1**). We defined a set of sequential treatment phases, with transitions based on clinically significant therapeutic events, enabling time-to-event analysis mapped onto "time-to-phase" transitions.

Phase I (Initial Management):

Upon clinical diagnosis of active PsA, patients are evaluated for the presence of poor prognostic factors. If absent, non-steroidal anti-inflammatory drugs (NSAIDs) or local

corticosteroid injections are recommended. If poor prognostic features are present, or if disease activity persists despite NSAIDs, escalation to Phase II is initiated.
Phase II (csDMARD Initiation):
Patients enter Phase II upon initiation of a conventional synthetic disease-modifying antirheumatic drug (csDMARD), such as methotrexate (MTX), leflunomide, or sulfasalazine. Transition criterion: Start of MTX or other csDMARD. Improvement is assessed at 3 months and target attainment at 6 months.

Phase III (First b/tsDMARD):
Patients with inadequate response to csDMARDs are escalated to biological or targeted synthetic DMARDs (b/tsDMARDs), including TNF inhibitors, IL-17 inhibitors, or JAK inhibitors. Choice of agent is stratified by clinical phenotype (e.g., enthesitis, axial disease). Transition criterion: Start of bio or tsDMARD.

Phase IV (Switch within b/tsDMARD class):
Patients who do not improve after 3–6 months on their first b/tsDMARD switch to an alternative b/tsDMARD, either within the same class or to another mechanism of action. Transition criterion: Stop first bio or tsDMARD due to treatment failure.

Phase IV+ (Multiple b/tsDMARD Failures):
A further escalation phase is defined for patients failing multiple b/tsDMARDs. Transition criterion: Stop two or more b/tsDMARDs due to ineffectiveness. This phase reflects advanced refractory disease requiring repeated therapeutic switching. Note*:* Transitions are only modeled when enough data points are available in the patient cohorts.

Clustering (k-means)
To capture heterogeneous treatment trajectories over time, we constructed an 8-dimensional clustering space for each patient, representing both the occurrence and timing of entry into sequential medication phases (P2–P5). This space was defined as:

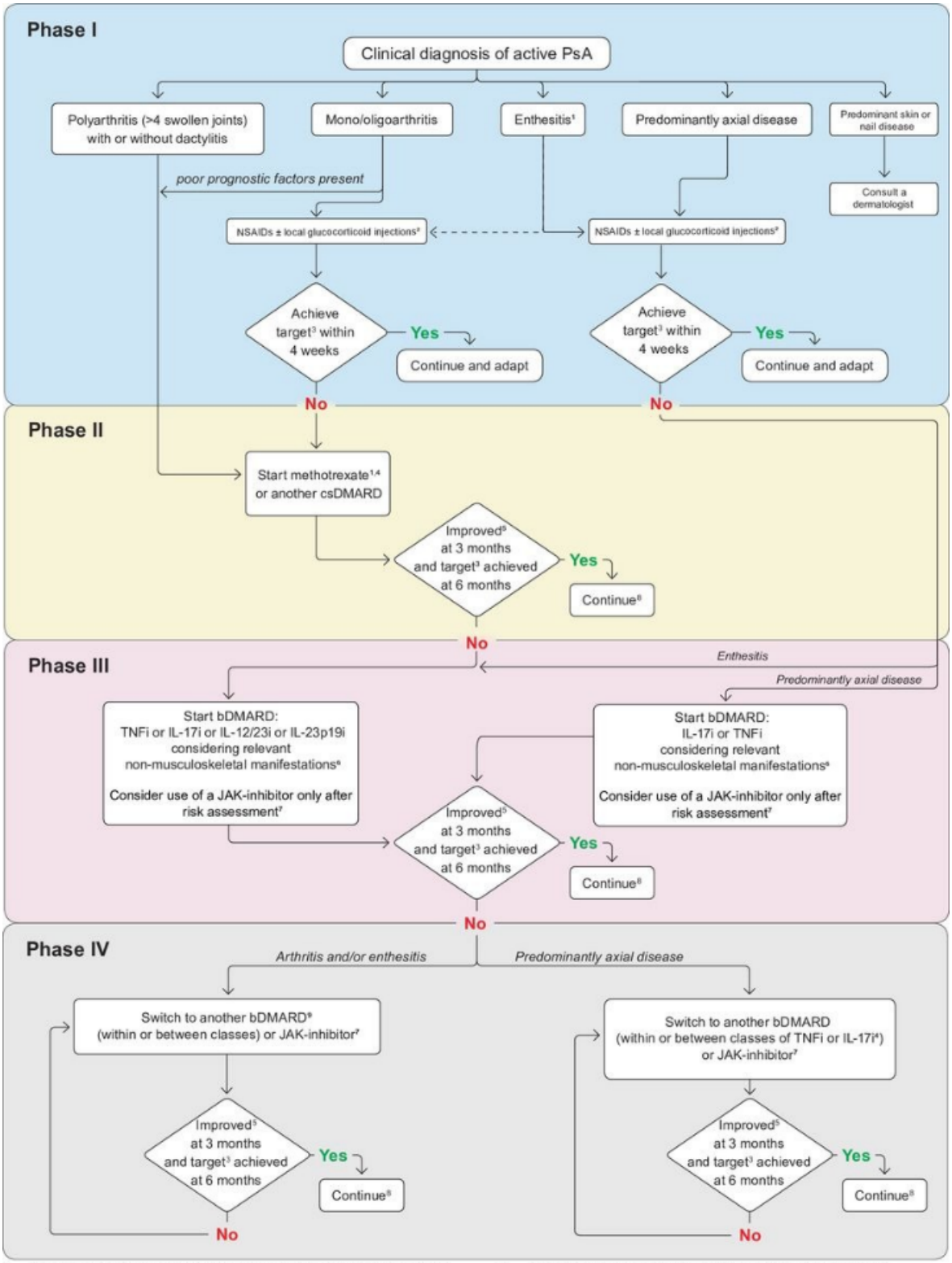
- Binary indicators event2, event3, event4, event5: each set to 1 if the patient reached the corresponding treatment phase during a defined 36-month observation period, and 0 otherwise.
- Continuous variables time2, time3, time4, time5: representing the number of days from the diagnosis date (or baseline visit) to the point of entering each respective phase.

For each phase, patients were evaluated for whether they met the criteria for treatment initiation. Time-to-event variables were calculated accordingly.

For patients who did not reach a given phase, the binary event indicator was set to 0, and the corresponding time-to-event was defined as the time from diagnosis to censoring (determined as the earliest of 36+delay months or end of observation). This allowed censored trajectories to be explicitly represented in the clustering space.

To ensure comparability across features, all eight variables were standardized using z-score normalization. Patients with missing values in any of the features were excluded, resulting in a cohort with complete longitudinal treatment data across all four phases.

We applied k-means clustering (with k=12) to this standardized feature space, effectively grouping patients based on the structure and tempo of their treatment progression over time.



1. *Some studies suggest that enthesitis may respond to methotrexate, but the level of evidence is low.*
2. *No glucocorticoids for axial disease.*
3. *The target is remission or low disease activity (especially with long standing disease) in accordance with the treat-to-target recommendations.*
4. *Preferred in the presence of relevant skin involvement, however in case of concomitant inflammatory bowel disease or uveitis, a TNF monoclonal antibody or (for IBD) IL-23i or 12/23i or JAKi is recommended.*
5. *Improvement means at least 50% reduction in disease activity.*
6. *Arthritis/enthesitis: TNFi or IL-17i or IL-12/23i or IL-23p19i; Skin: IL-17i or IL-12/23i or IL-23p19i; Uveitis: anti-TNF monoclonal antibody; IBD: anti-TNF monoclonal antibody or IL-12/23i or IL-23p19i or JAKi; Consider using PDE4i in mild disease if bDMARD and JAKi is inappropriate.*
7. *For JAK-inhibitors, caution is needed for patients aged 65 years or above, current or past long-time smokers, with a history of atherosclerotic cardiovascular disease or other cardiovascular risk factors or with other malignancy risk factors; with known risk factors for venous thromboembolism.*
8. *Consider tapering in sustained remission.*
9. *Including abatacept.*

Figure 1 PsA Treatment Escalation phases from EULAR guidelines (Gossec et al., 2024). The phases are used to define events in COX survival analysis, to analyze treatment persistence.

Each resulting cluster reflects a distinct trajectory pattern — for instance, early treatment escalation, delayed phase transition, or stagnation in earlier phases.

After clustering, we analyzed each group's clinical composition using two independent axes of stratification:

1. Terminal phase: For each patient, we identified the final treatment phase reached during the follow-up window. This enabled us to quantify, within each cluster, the proportion of patients who remained in early phases versus those who progressed further.
2. Diagnosis timing category: Patients were also labeled based on the timing of their diagnosis relative to disease onset, categorized as *very early*, *early*, *intermediate*, or *late*. We used this stratification to evaluate whether certain clusters were enriched for specific diagnostic phenotypes, suggesting a relationship between initial presentation and long-term treatment trajectory.

Together, these analyses allowed us not only to define data-driven patient subgroups, but also to interpret them clinically — revealing how patterns of medication timing and progression are influenced by underlying diagnostic characteristics and early disease dynamics.

Survival analysis- COX model and kNN Feature Engineering

Time-to-Event Outcomes

To quantify how diagnosis timing impacts the probability of reaching higher EULAR treatment phases over time, we fitted Cox proportional hazards models predicting the time-to-event for reaching Phase 2, Phase 3, or Phase 4. Each Cox model was fit separately for the time-to-event. The observation window was capped at 36 months, and patients who had not experienced the event within that period (plus a 2-month grace period) were right-censored.

Covariates

To isolate the effect of diagnosis timing on treatment persistence while controlling for potential confounding factors, the following covariates were included in the model:

Baseline demographic variables:

- Age
- Sex
- BMI
- Smoking status

Clinical variables:

- Disease phenotype (subtype classification)

Diagnosis group:

- Patients were grouped into Very Early, Early, Mid, or Late diagnosis categories depending on how soon after symptom onset the diagnosis was made.

These covariates were chosen to control for disease presentation and baseline patient characteristics that could influence treatment persistence.

An engineered covariate based on k-nearest neighbor (kNN) similarity in clinical trajectories was also investigated.

Constructing the kNN Feature

To incorporate personalized trajectory-based risk into the survival models, we engineered an out-of-fold (OOF) feature using k-nearest neighbors (kNN) regression, estimating time-to-

event based on clinical similarity in covariates. The resulting risk score was used as an additional covariate in Cox proportional hazards models predicting progression to EULAR treatment Phases 2, 3, or 4. To construct a predictive risk score from the training data without introducing data leakage, we implemented an Out-Of-Fold (OOF) strategy. The latter is a technique used during model training to construct a predictive risk score without causing data leakage. By dividing the training dataset into multiple folds, a model is iteratively trained on all but one-fold and then used to predict outcomes for the excluded fold—the "out-of-fold" portion. This process ensures that each data point receives a prediction from a model that has never seen it during training, preserving the integrity of performance metrics. The aggregated out-of-fold predictions form a reliable risk score across the entire training set.

Confidence Intervals and Statistical Significance
We estimated hazard ratios (HRs) along with 95% or 99% confidence intervals (CIs) for each covariate. The CIs were calculated as:

$$\text{CI} = e^{\beta \pm Z_\alpha \cdot \text{SE}(\beta)}$$

Where:

- β is the Cox model coefficient
- $SE(\beta)$ is the standard error of β
- Z_a is the Z-score corresponding to the confidence level (e.g., 1.96 for 95% and 2.576 for 99%)

P-values were computed using the Wald test, according to:

$$p = 2 \cdot \left(1 - \Phi(|\beta/\text{SE}(\beta)|)\right)$$

where Φ is the cumulative distribution function of the standard normal distribution.

Estimation of Model Uncertainty
To assess the stability and variability of the Cox model coefficients, we used bootstrapping:

- The model was refitted 1,000 times on resampled datasets (with replacement).
- For each bootstrap iteration, the standard errors and covariance matrices were recalculated to estimate the uncertainty in coefficient estimates.

This bootstrapped error estimation approach ensures robust inference by accounting for sampling variability in the estimation process.

2.3 Results

2.3.1 Characteristics of PsA patients

A total of 936 patients from the DEPAR cohort and 753 from Reuma.pt were included (**Table 4**), all with a baseline visit (T0) within the first year of PsA diagnosis. Age at diagnosis was similar between cohorts (median [IQR]: 51 [39–60] in DEPAR vs. 50 [39–59] in Reuma.pt; p = 0.215). Sex distribution differed significantly, with more males in Reuma.pt (56.2% vs. 48.7%; p = 0.003). Ethnicity data were not directly comparable due to different classification methods **Table 4**; **Appendix Table 2**). All DEPAR patients were Dutch; in Reuma.pt, 65.9% were recorded as White European, 2.4% as other ethnic backgrounds, and 31.7% as unknown.

Education levels appeared higher in DEPAR, with over half reaching ISCED levels 5–6 compared to 7.5% in Reuma.pt, although data were largely missing in Reuma.pt (94.7%) (**Table 4**; **Appendix Table 1**). Employment status among those below retirement age (<68 years for DEPAR, <66 for Reuma.pt) was similar (75.7% vs. 80.0%; p = 0.598), but missingness was high in Reuma.pt (92.6%).

Table 4 Baseline demographic characteristics of PsA patients in DEPAR and Reuma.pt cohorts.[1]

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
Age	51 [39-60]	50 [39-59]	0%	0%	0.215
Sex (male)	48.72%	56.18%	0%	0%	0.003
Ethnicity	-	-	0%	0%	NA
Dutch	100%	NA	-	-	-
White EU	NA	65.87%	-	-	-
Unknown	NA	31.74%	-	-	-
Other	NA	2.4%	-	-	-
Education (ISCED level)	-	-	13.25%	94.69%	NA
Level 1	4.31%	25%	-	-	-
Level 2	15.39%	12.5%	-	-	-
Level 3	14.04%	32.5%	-	-	-
Level 4	15.39%	22.5%	-	-	-
Level 5	31.89%	7.5%	-	-	-
Level 6	18.97%	0%			
Employment status	75.65%	80%	4.26%	92.57%	0.598

Medical history characteristics are shown in **Table 5**. Psoriasis was reported in 88.1% of DEPAR patients and 63.8% in Reuma.pt, but the duration before PsA onset was comparable (10 [4–22] vs. 10 [3–21] years; p = 0.308). HLA-B27 positivity was similar (15.4% vs. 13.4%; p = 0.470), though with high missingness (61% in DEPAR, 41.3% in Reuma.pt). BMI and obesity rates were also comparable (BMI: 27.7 [24.8–31.3] vs. 28.2 [24.6–31.6]; p = 0.780; obesity: 33.5% vs. 32.1%; p = 0.900), though BMI data were missing in 89.6% of Reuma.pt.

Smoking history (current or past) was more frequently reported in DEPAR (64.1% vs. 40.4%; p < 0.001), while among smokers, pack year exposure was higher in Reuma.pt (25 [10–40] vs. 14 [6–25]; p < 0.001). Comorbidity prevalence was similar between cohorts (**Figure 2**), with cardiovascular disease present in 18.8% (DEPAR) and 18.2% (Reuma.pt), uveitis in 2.6% and 1.9%, and IBD in 0.5% in both. Statistical comparisons were not conducted for these due to the lack of negative reporting.

Table 5 Medical history at baseline (T0) in DEPAR and Reuma.pt [1]

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
PsO diagnosis[2]	88.14%	63.75%	NA	NA	NA

[1] Continuous variables are summarized as medians with interquartile ranges [IQR] and compared using the Kruskal-Wallis test. Categorical variables are presented as percentages and compared using the Chi-square test.
[2] Comorbidities and psoriasis diagnosis were recorded only when present; absence of reporting does not confirm absence of disease. No statistical comparisons were performed.

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
PsO before PsA onset (years)	10 [4-22]	10 [3-21]	0%	0%	0.308
HLA-B27 (+)	15.38%	13.35%	61%	41.3%	0.470
BMI	27.73 [24.76-31.25]	28.20 [24.62-31.59]	14.1%	89.64%	0.78
Obesity	33.46%	32.05%	14.1%	89.64%	0.9
Smoking history (yes)	64.11%	40.42%	13.57%	50.2%	<0.001
Packyears	14 [6-25]	25 [10-40]	0%	52.3%	<0.001
Comorbidities[2]					
CVD	18.8%	18.19%	NA	NA	NA
Uveitis	2.56%	1.86%	NA	NA	NA
IBD	0.53%	0.53%	NA	NA	NA

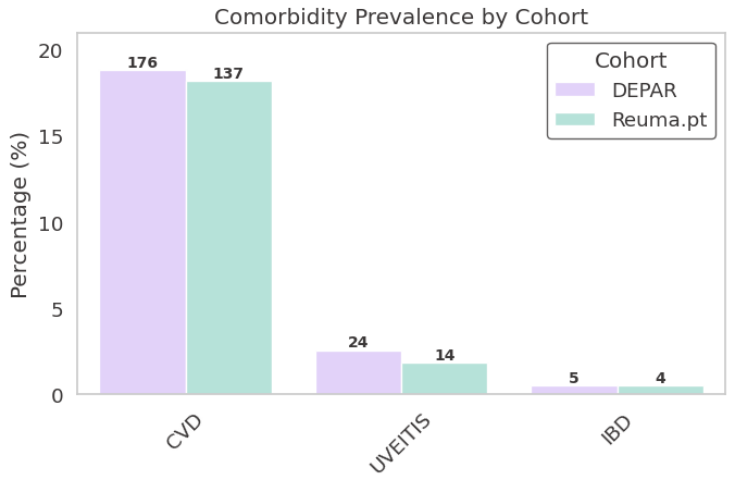


Figure 2 Prevalence of comorbidities in DEPAR and Reuma.pt at baseline (T0).

Disease phenotypes at baseline are presented in **Table 6** and **Figure 3**. Oligoarthritis was most common in both cohorts (34.9% DEPAR, 35.8% Reuma.pt). Polyarthritis was more frequent in Reuma.pt (41.2% vs. 25.8%), while DEPAR showed more monoarthritis (17.7% vs. 0.1%). Axial disease was reported more often in Reuma.pt (21.0% vs. 3.0%), whereas DEPAR had higher rates of dactylitis and enthesitis (10.7% and 8.0% vs. 1.8% and 0.1%). The distribution of phenotypes differed significantly between cohorts ($p < 0.001$), with lower representation of severe phenotypes in DEPAR likely reflecting differences in recruitment strategies. Reuma.pt included patients initiating biologic therapy, potentially leading to overrepresentation of axial and polyarticular disease.

Table 6 Disease phenotype distribution at baseline (T0) in DEPAR and Reuma.pt.[1]

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
Disease phenotype	-	-	3.21%	1.33%	<0.001
Axial disease	2.98%	21%	-	-	-
Dactylitis	10.71%	1.75%	-	-	-
Enthesitis	7.95%	0.13%	-	-	-
Monoarthritis	17.66%	0.13%	-	-	-
Oligoarthritis	34.88%	35.8%	-	-	-
Polyarthritis	25.83%	41.18%	-	-	-

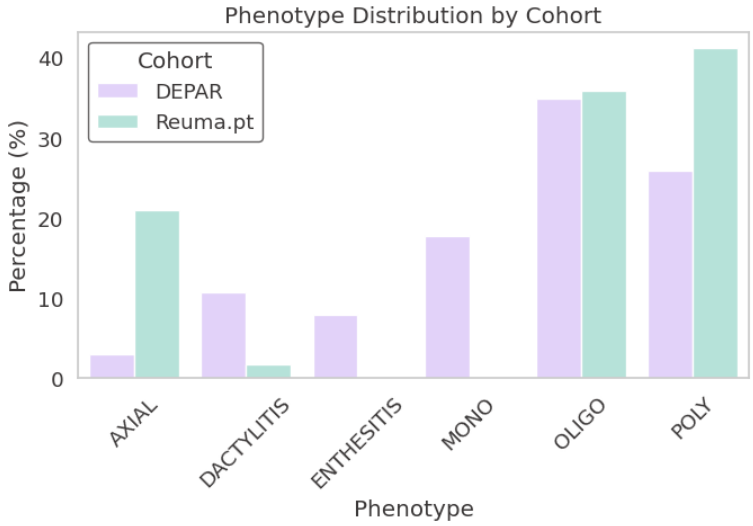


Figure 3 Distribution of disease phenotypes in DEPAR and Reuma.pt at baseline (T0).

Clinical and disease activity measures are summarized in **Table 7**. Reuma.pt patients had higher joint counts (SJC66: 3 [0–6], TJC68: 4 [1–8]) than DEPAR (SJC66: 1 [0–3], TJC68: 2 [0–6]; $p < 0.001$). Dactylitis count and PASI scores were also higher ($p < 0.001$), although missingness in Reuma.pt exceeded 88%. LEI scores were slightly higher in DEPAR (0 [0–1] vs. 0 [0–0]; $p < 0.001$). Symptom duration was longer in Reuma.pt (1.17 vs. 0.81 years; $p < 0.001$), while time since physician diagnosis was shorter (0.27 vs. 0.34 years; $p < 0.001$). DAPSA scores were higher in Reuma.pt (22.8 [12.9–34.5] vs. 14.3 [7.7–21.7]; $p < 0.001$), and fewer patients reached MDA (2.9% vs. 24.9%; $p < 0.001$). CRP levels were lower in Reuma.pt (1.1 vs. 3.0 mg/L; $p < 0.001$).

Table 7 Clinical assessment and disease activity at baseline (T0) in DEPAR and Reuma.pt.[1]

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
Clinical assessment					
SJC66	1 [0-3]	3 [0-6]	1.4%	24.7%	<0.001

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
Clinical assessment					
TJC68	2 [0-6]	4 [1-8]	1.4%	24.7%	<0.001
LEI	0 [0-1]	0 [0-0]	0.96%	66.27%	<0.001
PASI	1.2 [0-3.3]	3.1 [0.75-8.6]	29.59%	89.51%	<0.001
Dactylitis count	0 [0-0]	0 [0-2]	1.07%	88.85%	<0.001
Symptom onset (patient) (years)	0.81 [0.29-2.66]	1.17 [0.42-3.75]	1.28%	2.13%	<0.001
Disease duration-physician PsA diagnosis (years)*	0.34 [0.1-1.74]	0.27 [0.04-0.54] *	1.28%	0%	<0.001
Disease activity					
DAPSA	14.3 [7.71-21.7]	22.83 [12.93-34.46]	26.4%	68.92%	<0.001
PASDAS	3.86 [2.91-4.87]	6.56 [6.47-7.40]	35.36%	99.2%	NA
MDA	24.91%	2.86%	11.65%	21.12%	<0.001
CRP (mg/L)	3 [0-9]	1.1 [0.31-2.97]	13.78%	33.86%	<0.001

* In Reuma.pt, newly diagnosed patients are defined as those whose baseline visit occurs within one year of their diagnosis. In patients where only the year of diagnosis was available, July 1st was used as an estimated diagnosis date, which means the actual timing could be off by up to six months in either direction.

Patient-reported outcomes are shown in **Table 8**. VAS pain was higher in Reuma.pt (50 [29–70] vs. 43.5 [21–66]; p = 0.007), while VAS global was similar (p = 0.512). HAQ scores were slightly higher in Reuma.pt (0.75 [0.12–1.38] vs. 0.62 [0.25–1]; p = 0.094). HADS scores indicated greater anxiety and depression in Reuma.pt (anxiety: 8 [4–11] vs. 4 [2–7], p < 0.001; depression: 6 [1–9] vs. 3 [1–6], p = 0.05), although data were missing for 95.1% of patients in Reuma.pt.

Table 8 Patient-reported outcomes (PROs) at baseline in DEPAR and Reuma.pt.[1]

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
HAQ	0.62 [0.25-1]	0.75 [0.12-1.38]	14.53%	75.7%	0.094
SF36RE	75 [50-100]	33.33 [0-83.33]	20.19%	94.42%	<0.001
SF36BP	51 [41-62]	36.5 [22-51]	14%	94.16%	<0.001
SF36RP	50 [37.5-81.25]	25 [0-50]	14%	94.3%	<0.001
SF36GH	55 [40-67]	45 [39.25-55.5]	14%	94.2%	0.013
SF36VI	56.25 [37.5-68.75]	45 [30.62-50]	20.19%	94.3%	0.004
SF36MH	75 [60-85]	56.4 [48.2-75]	20.19%	94.3%	<0.001
SF36SF	75 [62.5-100]	62.5 [37.5-87.5]	20.19%	94.2%	0.004

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
SF36PF	70 [50-85]	45 [25-76.25]	14%	94.2%	0.001
VAS pain	43.5 [21-66]	50 [29-70]	20.94%	72.64%	0.007
VAS global	44 [19-64]	41 [20-60]	19.55%	58.3%	0.512
HADS anxiety	4 [2-7]	8 [4-11]	54.27%	95.09%	<0.001
HADS depression	3 [1-6]	6 [1-9]	54.27%	95.09%	0.05
GRACE	2.88 [1.68-4.1]	NA	44.12%	100%	NA

Baseline treatment patterns are reported in **Table 9**. Fewer patients were untreated in Reuma.pt (20%) compared to DEPAR (43%). Use of biologic DMARDs (19% vs. 3%) and glucocorticoids (28.6% vs. 7.8%) was higher in Reuma.pt. Methotrexate use was comparable (49.2% vs. 55.1%), while other csDMARDs were more frequently used in Reuma.pt (15.3% vs. 5.3%). Use of tsDMARDs was rare in both cohorts.

Highlight

A total of 936 patients from the DEPAR cohort and 753 from Reuma.pt were included. DEPAR and Reuma.pt cohorts showed differences in baseline characteristics including sex, smoking, PsA phenotype, TJC and SJC, LEI, PASI, dactylitis, disease duration, symptom onset, DAPSA, PASDAS, MDA, pain, and HADS anxiety/ depression, and in arthritis treatment. On the other hand, age, employment, PsO prior to PsA, BMI, comorbidities/manifestations, VAS global and HAQ were comparable between the two cohorts.

Table 9 Treatment at baseline (T0) in DEPAR and Reuma.pt.[1]

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
Medication	-	-	NA	NA	NA
None	43%	20%	-	-	-
Glucocorticoids	7.8%	28.6%	-	-	-
Methotrexate	49.2%	55.1%	-	-	-
csDMARDs	5.3%	15.3%	-	-	-
tsDMARDs	0.1%	0%	-	-	-
bDMARDs	3%	19%	-	-	-

2.3.2 Outcomes on sex differences

A total of 810 female and 879 male patients were included in the sex-based baseline analysis (**Table 10**). All participants had a baseline visit within one year of PsA diagnosis. Male patients were older at baseline than female patients (median [IQR]: 51 [41–60] vs. 49 [36–58] years; $p < 0.001$). Ethnicity distributions were similar between sexes, though a higher proportion of female patients were recorded as Dutch (59.3% vs. 51.9%) (**Table 10**; **Appendix Table 3**).

Educational levels did not differ significantly between sexes (p = 0.094), although female participants were more frequently represented in higher ISCED categories (e.g., level 5: 34.5% vs. 27.2%). Employment status was significantly higher among males (80.7%) compared to females (71.3%; p = 0.002).

Table 10 Baseline (T0) demographic characteristics of female and male PsA patients.[1]

	T0		
Outcomes	**Female (N=810)**	**Male (N=879)**	**p-value**
Age	49 [36-58]	51 [41-60]	<0.001
Missing %	0%	0%	-
Ethnicity	-	-	NA
Missing %	0%	0%	-
Dutch	59.26%	51.88%	-
White EU	28.02%	30.6%	-
Unknown	11.98%	16.15%	-
Other	0.75%	1.36%	-
Education (ISCED)	-	-	0.094
Missing %	48.39%	50.63%	
Level 1	4.07%	6.45%	-
Level 2	15.55%	14.98%	-
Level 3	12.68%	17.05%	-
Level 4	16.27%	15.21%	-
Level 5	34.45%	27.19%	-
Level 6	16.99%	19.12%	-
Employment status	71.3%	80.71%	0.002
Missing %	40.62%	46.99%	-

Medical history characteristics by sex are shown in **Table 11**. Psoriasis was more frequently reported in males (80.0%) than in females (74.3%), though the duration of psoriasis before PsA onset was similar (p = 0.326). HLA-B27 positivity did not differ between sexes (14.6% in females vs. 13.9% in males; p = 0.863). Obesity was more prevalent in females (37.8% vs. 28.7%; p = 0.005), while a history of smoking was more frequently reported in males (60.0% vs. 53.2%; p = 0.021). Among patients with smoking history, cumulative exposure (pack years) was comparable between sexes (p = 0.050).

Table 11 Medical history of female and male PsA patients at baseline (T0).[1]

	T0		
Outcomes	**Female (N=810)**	**Male (N=879)**	**p-value**
PsO diagnosis[2]	74.32%	79.98%	NA
Missing %	NA	NA	-
PsO before PsA onset (years)	9.99 [2.96-21.74]	10.19 [3.93-21.56]	0.326

	T0		
Outcomes	**Female (N=810)**	**Male (N=879)**	**p-value**
Missing %	0%	0%	-
HLA-B27 (+)	14.62%	13.94%	0.863
Missing %	51.85%	52.67%	-
BMI	28.23 [24.49-31.91]	27.46 [24.96-31.02]	0.445
Missing %	44.07%	51.19%	-
Obesity	37.75%	28.67%	0.005
Missing %	44.07%	51.19%	-
Smoking history (yes)	53.17%	60%	0.021
Missing %	29.88%	30.03%	-
Packyears	14 14.74 [5.26-25.26]	15.42 [7.89-31.32]	0.05
Missing %	8.28%	8,61%	-

Comorbidity data at baseline (T0) and over the three-year period (T0-T36) are shown in **Table 12** and **Figure 4**. Cardiovascular disease (CVD) was similarly reported in both sexes at baseline (18.6% in females vs. 18.4% in males) and remained comparable over time (30.1% vs. 30.6%). Uveitis was more frequently reported in females both at baseline (2.6% vs. 1.9%) and over three years (5.2% vs. 2.8%). Inflammatory bowel disease (IBD) was infrequently reported in both sexes.

Table 12 Comorbidities reported in female and male PsA patients at baseline (T0) and during the 3-year follow-up period (T0-T36).[1,2]

	T0		T0-T36	
Outcomes	**Female (N=810)**	**Male (N=879)**	**Female (N=810)**	**Male (N=879)**
Comorbidities	-	-	-	-
Missing %	NA	NA	NA	NA
CVD	18.64%	18.43%	30.12%	30.6%
Uveitis	2.59%	1.9%	5.19%	2.84%
IDB	0.74%	0.34%	1.24%	0.57%

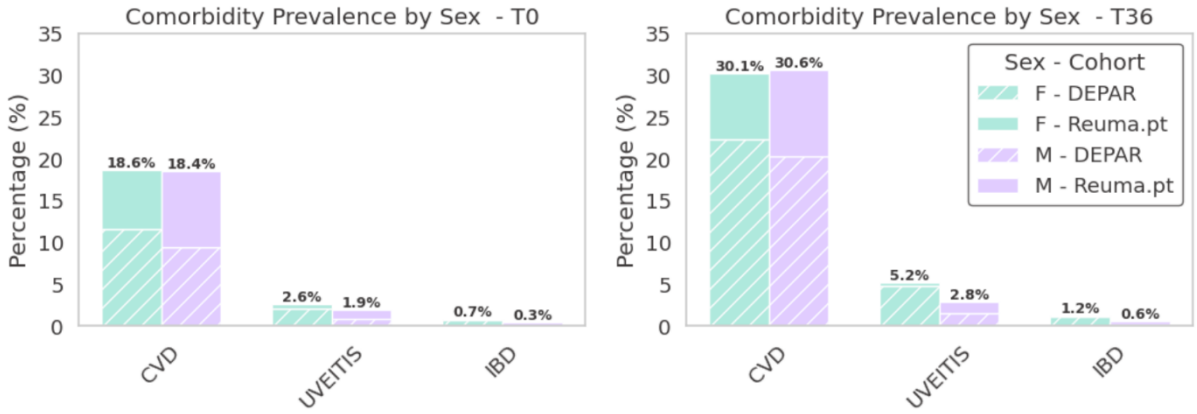


Figure 4 Prevalence of comorbidities by sex at baseline (T0) and during the 3-year follow-up period (T36).

Disease phenotype distribution at baseline (T0) is presented in **Table 13** and **Figure 5**. The overall distribution differed significantly between sexes ($p < 0.001$). Axial disease was more common in males (12.5%) than females (9.6%), while enthesitis and monoarthritis were more frequently reported in females (6.6% and 11.1%) compared to males (2.4% and 8.6%). Dactylitis was more frequent in males (7.3% vs. 6.0%). Oligoarthritis (34.0% in females vs. 36.5% in males) and polyarthritis (32.7% in both groups) were the most common phenotypes in both sexes, with similar frequencies observed.

Table 13 Baseline (T0) distribution of PsA phenotypes in female and male patients.[1]

	T0		
Outcomes	**Female (N=810)**	**Male (N=879)**	**p-value**
Disease phenotype	-	-	<0.001
Missing %	3.09%	1.71%	-
Axial disease	9.55%	12.5%	-
Dactylitis	5.99%	7.29%	-
Enthesitis	6.62%	2.43%	-
Monoarthritis	11.08%	8.56%	-
Oligoarthritis	34.01%	36.46%	-
Polyarthritis	32.74%	32.75%	-

Clinical assessment and disease activity outcomes by sex at baseline and after 36 months are presented in **Table 14**. Females reported a slightly longer symptom duration before diagnosis (1.02 vs. 1.00 years; $p = 0.05$). TJC68 were consistently higher in females than males at both baseline (3 [1–8] vs. 2 [0–6]; $p < 0.001$) and follow-up (0 [0–3] vs. 0 [0–1]; $p < 0.001$), while SJC66 were similar. LEI scores were higher in females at both time points ($p < 0.001$). Males had higher PASI scores at baseline (2 [0.4–4.4] vs. 0.8 [0–2.8]; $p < 0.001$) and follow-up ($p = 0.005$).

Disease activity was higher in females, with higher DAPSA scores at baseline (17.0 vs. 15.2; $p = 0.005$) and follow-up (9.9 vs. 5.7; $p < 0.001$). Minimal Disease Activity (MDA) was achieved less frequently in females at both time points (baseline: 13.3% vs. 17.9%, $p = 0.022$; follow-up: 32.8% vs. 50.5%, $p < 0.001$). CRP levels were similar across sexes.

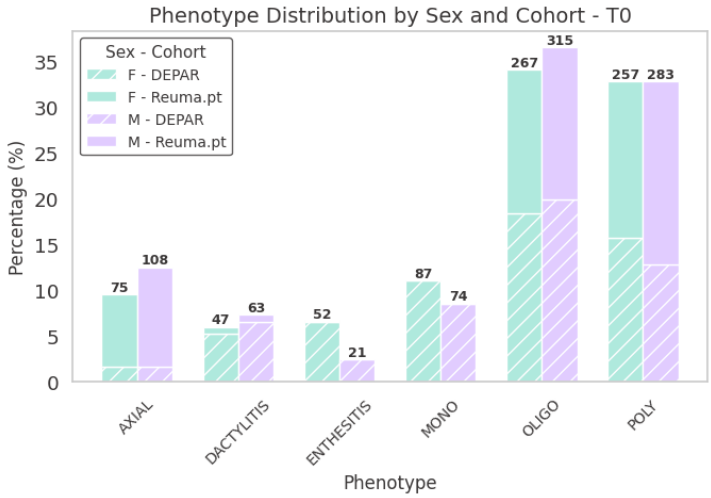


Figure 5 Distribution of PsA phenotypes by sex at baseline (T0).

Table 14 Clinical assessment and disease activity in female and male PsA patients at baseline (T0) and 3-year follow-up (T36).[1]

	T0			**T36**		
Outcomes	**Female (N=810)**	**Male (N=879)**	**p-value**	**Female (N=349)**	**Male (N=375)**	**p-value**
Clinical assessment						
SJC66	1 [0-5]	2 [0-5]	0.15	0 [0-1]	0 [0-1]	0.068
Missing %	11.23%	12.29%	-	3.72%	6.13%	-
TJC68	3 [1-8]	2 [0-6]	<0.001	0 [0-3]	0 [0-1]	<0.001
Missing %	11.23%	12.29%	-	3.72%	6.13%	-
LEI	0 [0-1]	0 [0-0]	<0.001	0 [0-1]	0 [0-0]	<0.001
Missing %	27.28%	32.65%	-	20.06%	26.67%	-
PASI	0.8 [0-2.8]	2 [0.4-4.4]	<0.001	0 [0-1]	0.1 [0-2.38]	0.005
Missing %	53.09%	59.27%	-	50.14%	54.67%	-
Dactylitis count	0 [0-0]	0 [0-0]	0.981	0 [0-0]	0 [0-0]	0.933
Missing %	36.42%	43.69%	-	28.94%	36.53%	-
Symptom onset (pat.)	1.02 [0.38-3.17]	1 [0.29-2.89]	0.05	NA	NA	NA
Missing %	1.6%	1.71%	-	NA	NA	-
Physician PsA diagn.	0.38 [0.1-0.96]	0.25 [0.07-0.61]	<0.001	NA	NA	NA
Missing %	0.62%	0.8%	-	NA	NA	-
Disease activity						
DAPSA	17 [10.53-25.48]	15.2 [7.5-24]	0.005	9.9 [4.97-15.7]	5.7 [2-11.7]	<0.001
Missing %	45.93%	44.82%	-	38.11%	38.93%	-
PASDAS	4.04 [3.07-4.87]	3.79 [2.74-4.93]	0.09	3.15 [2.16-4.20]	2.32 [1.52-3.49]	<0.001

Missing %	63.58%	64.05%	-	52.15%	58.4%	-
MDA	13.29%	17.88%	0.022	32.78%	50.46%	<0.001
Missing %	16.42%	15.36%	-	13.47%	13.33%	-
CRP	2 [0.25-6]	2 [0.28-6.54]	0.78	1 [0-4]	0.53 [0.05-3.5]	0.367
Missing %	22.96%	22.53%	-	17.19%	20.8%	-

PROs at baseline (T0) and after 36 months (T36) are shown in **Table 15**. Across both time points, females reported significantly higher levels of pain, global disease perception, disability, and psychological burden compared to males.

At baseline, HAQ scores were higher in females (0.75 [0.38–1.25]) than in males (0.50 [0.12–1.00]; $p < 0.001$), and this difference persisted at follow-up (0.75 vs. 0.12; $p < 0.001$). VAS pain and global scores were also consistently higher in females (baseline pain: 51 vs. 39; follow-up: 32.5 vs. 20; $p < 0.001$). HADS scores indicated greater anxiety and depressive symptoms in females at both baseline and follow-up ($p < 0.01$). Although both sexes showed improvement over time in all PROs, differences between groups remained significant.

Table 15 PROs in female and male PsA patients at baseline (T0) and 3-year follow-up (T36).

	T0			**T36**		
Outcomes	**Female (N=810)**	**Male (N=879)**	**p-value**	**Female (N=349)**	**Male (N=375)**	**p-value**
HAQ	0.75 [0.38-1.25]	0.5 [0.12-1]	<0.001	0.75 [0.25-1.25]	0.12 [0-0.62]	<0.001
Missing %	39.51%	43.91%	-	27.51%	28.8%	-
SF36RE	75 [50-100]	83.33 [58.33-100]	<0.001	75 [50-100]	100 [66.67-100]	<0.001
Missing %	50.86%	55.52%	-	35.24%	38.4%	-
SF36BP	41 [32-62]	52 [41-72]	<0.001	52 [41-62]	62 [51-84]	<0.001
Missing %	47.41%	51.88%	-	34.1%	38.4%	-
SF36RP	50 [25-68.75]	62.5 [43.75-92.19]	<0.001	56.25 [43.75-75]	75 [56.25-100]	<0.001
Missing %	47.41%	51.99%	-	33.81%	38.4%	-
SF36GH	50 [35-65]	57 [42-72]	<0.001	52 [35-67]	62 [47-74.5]	<0.001
Missing %	47.41%	51.88%	-	33.81%	38.4%	-
SF36VI	50 [31.25-62.5]	56.25 [43.75-75]	<0.001	50 [37.5-62.5]	62.5 [50-81.25]	<0.001
Missing %	50.62%	55.63%	-	34.96%	38.4%	-
SF36MH	70 [55-85]	80 [60-90]	<0.001	75 [60-85]	85 [70-90]	<0.001
Missing %	50.62%	55.63%	-	34.96%	38.4%	-
SF36SF	62.5 [50-87.5]	75 [62.5-100]	<0.001	75 [62.5-100]	87.5 [62.5-100]	<0.001
Missing %	50.62%	55.52%	-	34.96%	38.4%	-
SF36PF	65 [45-80]	75 [50-90]	<0.001	65 [45-85]	90 [65-95]	<0.001

Missing %	47.41%	51.88%	-	33.81%	38.13%	
VAS pain	51 [28.25-70.75]	39 [19.25-63]	<0.001	32.5 [13-56.25]	20 [5-41]	<0.001
Missing %	40.49%	41.07%	-	28.94%	27.73%	-
VAS global	49 [24-66]	40 [17-60]	<0.001	23.5 [10-50]	13 [4-30]	<0.001
Missing %	35.19%	35.38%	-	23.78%	23.2%	-
HADS anxiety	5 [3-9]	4 [1-6]	<0.001	4 [3-8]	3 [1-5]	<0.001
Missing %	71.11%	73.72%	-	67.62%	68.27%	-
HADS depression	3 [1-8]	2 [1-5]	0.002	2 [1-6]	2 [1-3]	0.007
Missing %	71.11%	73.72%	-	67.62%	68.27%	-
GRACE						
Missing %	67.16%	70.76%	-	58.74%	64.8%	-

Longitudinal disease activity trajectories are shown in **Figure 6**. DAPSA scores declined over time in both sexes, indicating improvement. However, females consistently reported higher disease activity across all visits. The proportion of patients achieving minimal disease activity (MDA) increased in both groups over time but remained significantly lower in females at all time points. This trend widened at later visits.

Mixed linear model results (**Figure 7**) (**Appendix Table 4**) confirmed that male sex was associated with significantly lower DAPSA scores over time (–2.67, $p < 0.001$), while the interaction between sex and time was not significant, suggesting parallel improvement rates. Patients from the Reuma.pt cohort had higher DAPSA scores (2.61, $p < 0.001$), suggesting higher disease activity at baseline and during follow-up.

In the generalized estimating equation model for MDA (**Figure 7**) (**Appendix Table 5**), male sex was associated with a higher likelihood of achieving MDA (OR ~ 1.61, $p < 0.001$). A significant interaction between sex and time ($p < 0.001$) indicated a faster rate of improvement in males. Cohort and age also influenced MDA outcomes, with lower MDA rates observed in Reuma.pt and older patients.

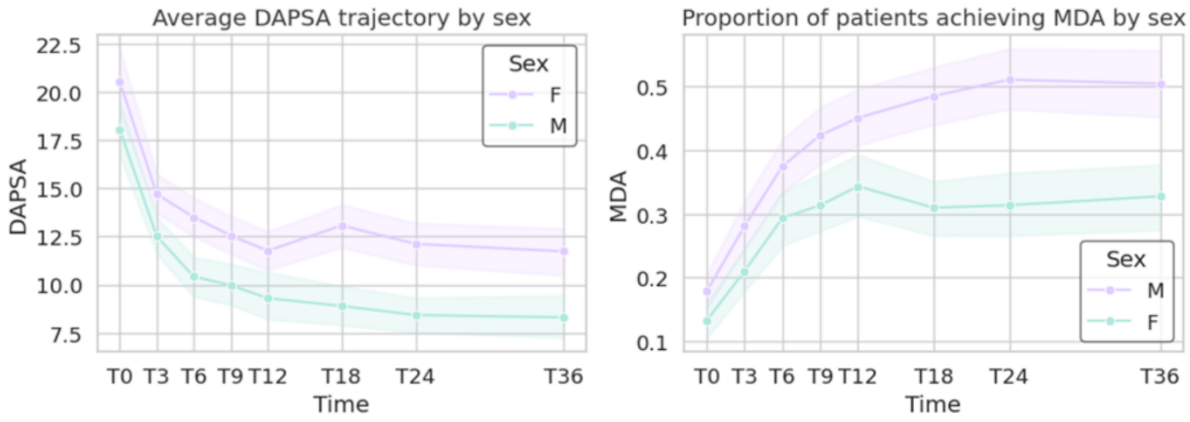


Figure 6 Left: Trajectory of average DAPSA scores over time by sex. Right: Proportion of PsA patients achieving Minimal Disease Activity (MDA) over time by sex. F: Female; M: Male.

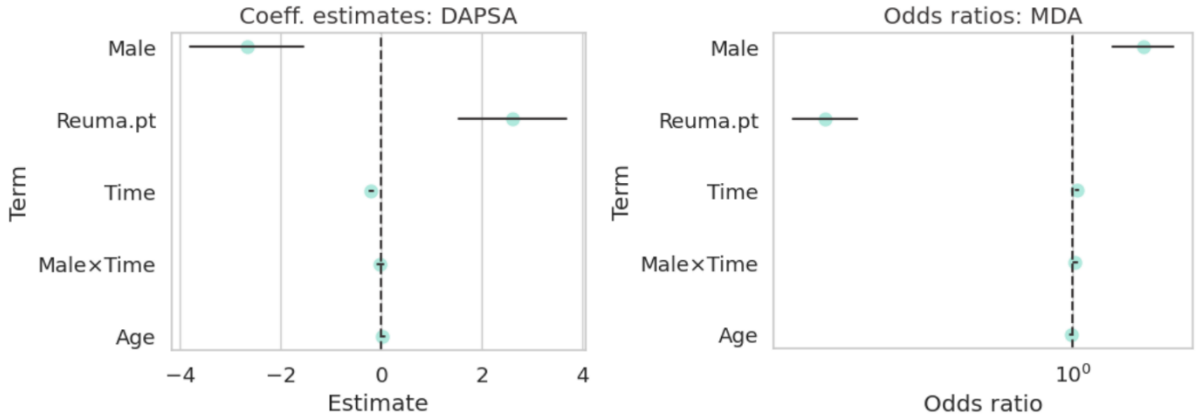


Figure 7 Left: Coefficient estimates from the mixed model assessing longitudinal DAPSA scores. Right: Odds ratios from the generalized estimating equations model for MDA.

Longitudinal trajectories for VAS-PAIN and HAQ are shown in **Figure 8**. Both sexes showed substantial improvement over time in pain and functional disability. However, women consistently reported higher scores across all visits, indicating greater perceived burden.

Mixed linear models (**Figure 9**) (**Appendix Table 6; Appendix Table 7**) confirmed that male sex was significantly associated with lower pain (–9.44, $p < 0.001$) and lower HAQ scores (0.278, $p < 0.001$). Time was a strong predictor of improvement in both outcomes ($p < 0.001$).

While the interaction between sex and time was not significant for VAS pain ($p = 0.463$), it was significant for HAQ (–0.003, $p < 0.001$), suggesting a slightly faster improvement in males. Age was associated with greater disability (0.007, $p < 0.001$), but not with pain. No significant effects of cohort were observed.

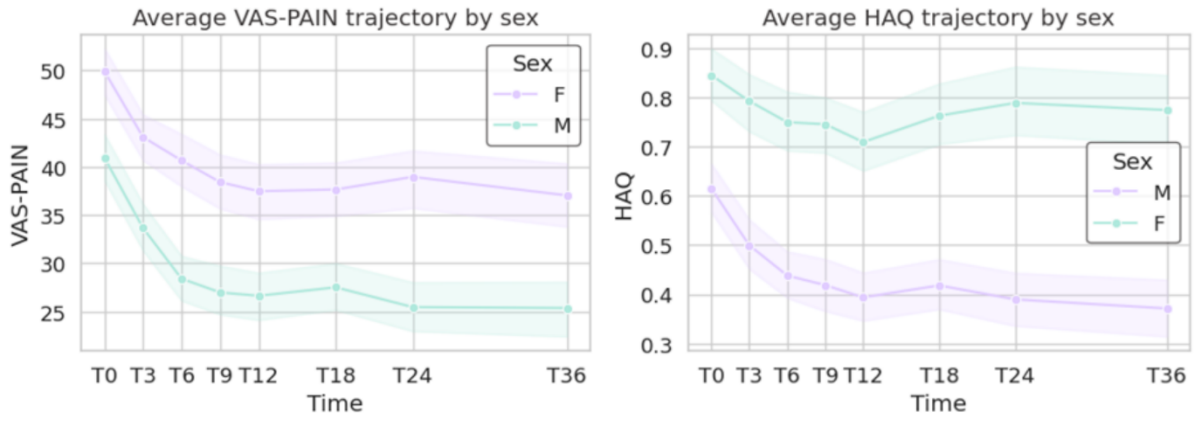


Figure 8 Left: Trajectory of average VAS pain scores over time by sex. Right: Trajectory of average HAQ scores over time by sex. F: Female; M: Male.

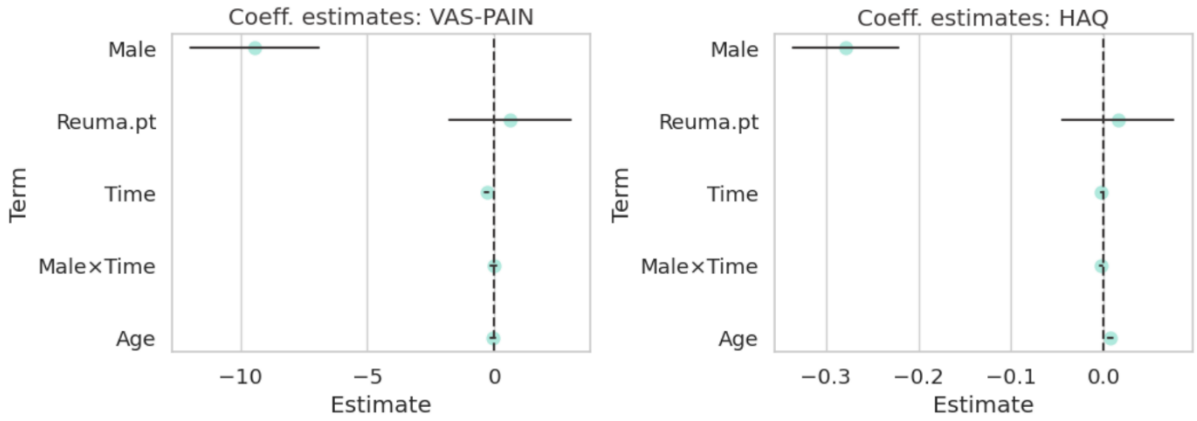


Figure 9 Left: Coefficient estimates from the mixed model assessing longitudinal VAS pain scores. Right: Coefficient estimates from the mixed model assessing longitudinal HAQ scores.

Treatment patterns over the 3-year period varied by medication and sex (**Figure 10**) (**Appendix Table 8**). At baseline, MTX was prescribed to 43% of males and 41% of females, while GC were more frequently used in females (15%) than males (13%). Use of bDMARDs increased from 8% at T0 to 31% at T24 in females and from 8% to 30% in males. In contrast, GC use declined, reaching 8% in both sexes by T36. MTX exposure peaked at T6–T9 in both sexes (up to 59% in males and 52% in females) but declined afterward, dropping to 49% in males and 37% in females by T36. csDMARD use remained relatively stable, with slightly higher usage in females (up to 19% at T18) compared to males (13% at T18). Only a small number of patients were prescribed tsDMARDs. The proportion of patients not receiving any treatment initially declined, reaching a minimum at T18 in males (25%) and at T12 in females (27%), before increasing again to 30% in males and 41% in females at T36.

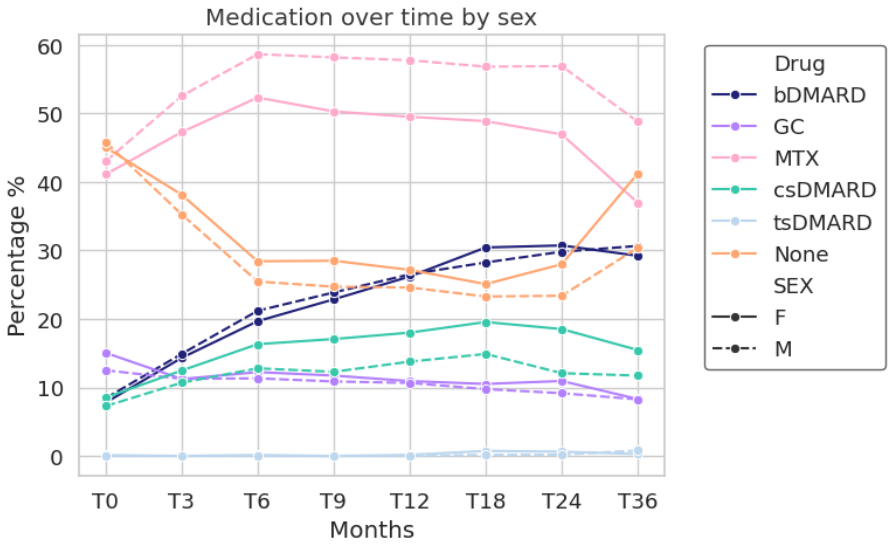


Figure 10 Medication class usage over time by sex. Percentages of patients receiving each medication class at each timepoint are shown separately for females (F) and males (M).

Over the 3-year follow-up, treatment duration differed by sex across several drug classes (**Figure 11**; **Table 16**). Males had longer median exposure than females to GC (234 vs. 120 days) and MTX (598 vs. 484 days). In contrast, cumulative exposure to csDMARDs was comparable between sexes (283 days).

Treatment retention over 36 months differed by sex across several medication classes (**Figure 12**) (**Appendix Table 9**). Males had significantly higher retention for MTX, bDMARDs, and GC, with differences apparent as early as 6 months and persisting through 36 months. For example, at 36 months, MTX retention was 51% in males vs. 41% in females, and bDMARD retention was 49% vs. 38%, respectively. GC retention was consistently lower in females across all time points (22% vs. 30% at 36 months). No significant sex differences were observed for csDMARDs. Log-rank tests confirmed statistically significant sex differences in retention for MTX ($p < 0.001$), bDMARDs ($p = 0.007$), and GC ($p = 0.001$).

Table 16 Summary of cumulative treatment duration by medication class and sex.[3]

	Female		Male	
Medication	**Median [IQR]**	**N**	**Median [IQR]**	**N**
GC	120 [42, 539]	253	234 [92, 633]	210
MTX	484 [177, 873]	607	598 [204, 1093]	666
csDMARD	283 [112, 637]	292	283 [92, 651]	216
bDMARD	500 [264, 821]	378	548 [289, 888]	339
tsDMARD	150 [37, 260]	20	162 [159, 181]	6

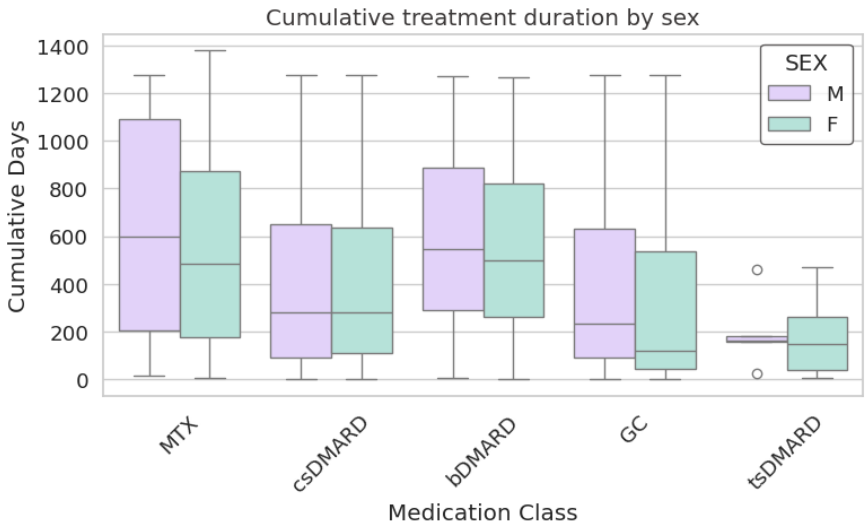


Figure 11 Median cumulative treatment duration (in days) over 3 years by drug class and sex. Boxes represent interquartile ranges. F: Female; M: Male.

Highlight

A total of 810 female and 879 male patients were included from DEPAR and Reuma.pt cohorts in the sex-based analysis. Compared to females, males were older, more likely to be employed and to smoke, and less likely to be obese or have uveitis. In contrast, females had higher tender joint counts and LEI, longer symptom and disease duration, higher DAPSA scores, greater VAS pain and global and HADS anxiety/depression, and were less likely to achieve MDA compared to males. Additionally, sex differences were observed in disease phenotype, and treatment duration and retention.

[3] Values represent median interquartile range [IQR] cumulative days on treatment. N indicates the number of patients receiving each treatment.

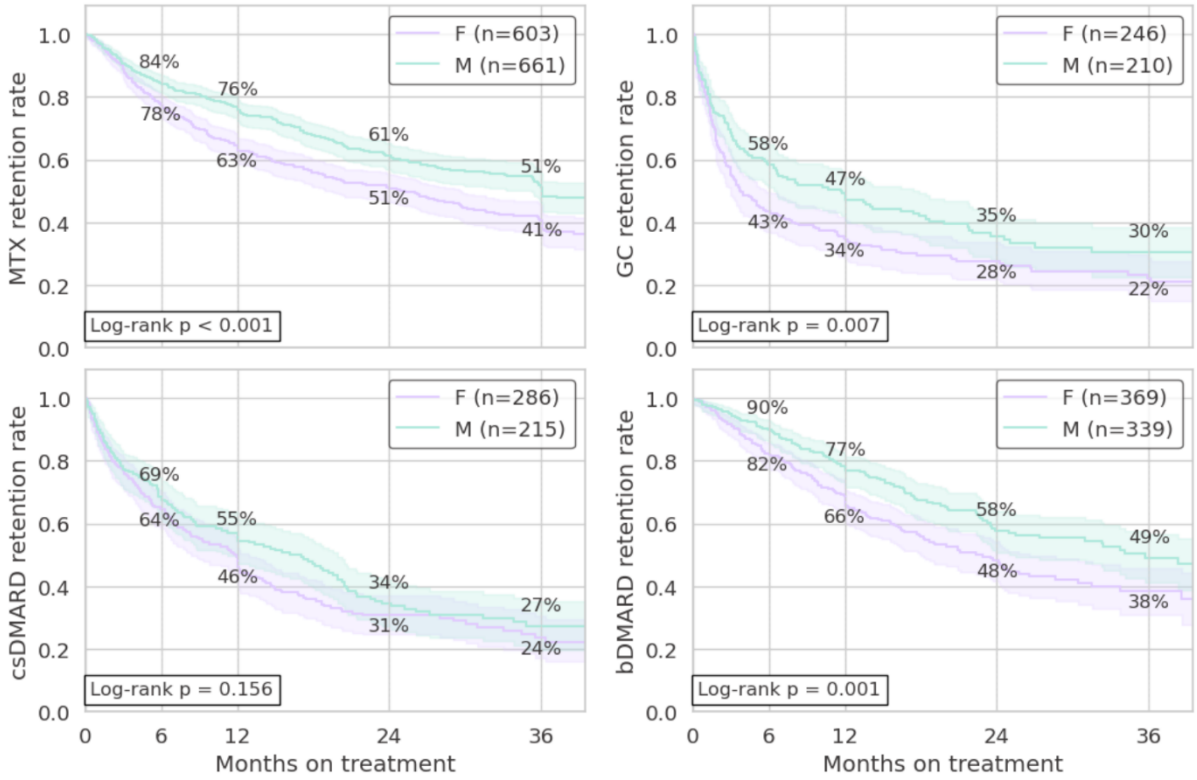


Figure 12 Kaplan-Meier curves of treatment retention by medication class and sex over 36 months. Curves display retention at 6, 12, 24, and 36 months along with sample sizes (n), with p-values from log-rank tests comparing retention between males (F) and females (F).

2.3.3 Outcomes on time to PsA diagnosis and disease prognosis

2.3.3.1 Data availability and population characteristics in DEPAR and Reuma.pt cohorts

DEPAR is an inception cohort, so newly diagnosed patients are registered in DEPAR cohort at times close to their diagnosis date and irrespective of disease activity. Patients in DEPAR represent a broad spectrum of disease trajectories, with no selection bias towards severe or mild disease presentation. Reuma.pt was initially established to register patients with rheumatic diseases receiving biological therapies, with the aim of monitoring treatment efficacy, safety, and long-term comorbidities (Canhão et al., 2011). Over time, the scope of Reuma.pt has broadened to include a wider range of treatments and patient profiles. Currently, approximately 50% of PsA patients registered in Reuma.pt have not been exposed to bDMARDs. Thus, the Reuma.pt cohort comprises more patients with PsA who are initiating biological treatment, thereby representing a population of more *difficult-to-treat* cases (this is also evident by the results in **Figure 7**). Unlike the DEPAR cohort, which includes a broader spectrum of PsA patients including also those with milder disease trajectories.

We associate time-to-PsA-diagnosis (i.e., early or late) with disease prognosis and treatment difficulty. For disease prognosis we use as outcomes MDA, PASDAS, DAPSA, GRACE and SF36 scores. MDA is available in both Reuma.PT and DEPAR cohorts, so we do the analysis on the combined population. PASDAS, DAPSA, GRACE and SF36 scores are highly missing in reuma.PT (see **Table 7**, **Table 8**), and the analysis is done on DEPAR population

Difficulty to treat outcome is the treatment phase according to EULAR guidelines. This analysis is done on DEPAR population, for two main reasons: i) to avoid introducing selection bias, as initially only patients initiating biologics were selected in this cohort this group is overrepresented, the inclusion of patients not treated with biologics in reuma.pt was progressive so we cannot use a simple cutoff year to filter out patients in the analysis, **and** ii)

we used as covariates in this analysis BMI and smoking exposure which are highly missing (see **Table 5**).

2.3.3.2 Differences in disease activity in Late vs Early Diagnosis groups: MDA

Posterior predictions revealed systematic differences in the probability of achieving MDA depending on the timing of diagnosis (**Figure 13**, left panel) (**Table 17**). Patients in the "Very Early" group had the highest mean predicted probability of MDA (mean: 0.388, 95% HDI: [0.371, 0.405]), while "Late" patients had the lowest (mean: 0.252, 95% HDI: [0.240, 0.263]). The intermediate groups ("Early" and "Intermediate") followed a gradual decline in probabilities, forming a consistent left-to-right gradient. This is further reflected in the posterior predictive distributions (**Figure 13**, right panel), which show clear, non-overlapping density curves for each group, emphasizing the probabilistic separation and lower likelihood of MDA among patients diagnosed later.

To assess temporal dynamics in treatment response, we evaluated posterior predictive probabilities of MDA at 6, 12, 24, and 36 months, stratified by diagnosis group (**Figure 14**). All groups showed increasing probabilities of achieving MDA over time, indicating general treatment effectiveness. However, the relative separation between groups persisted and even widened at later time points. By month 36, the probability of achieving MDA reached approximately 0.60 for "Very Early" and "Early" groups, compared to just over 0.50 for "Intermediate" and below 0.40 for the "Late" group. This consistent lag in MDA attainment among late-diagnosed patients suggests a lasting disadvantage in disease control.

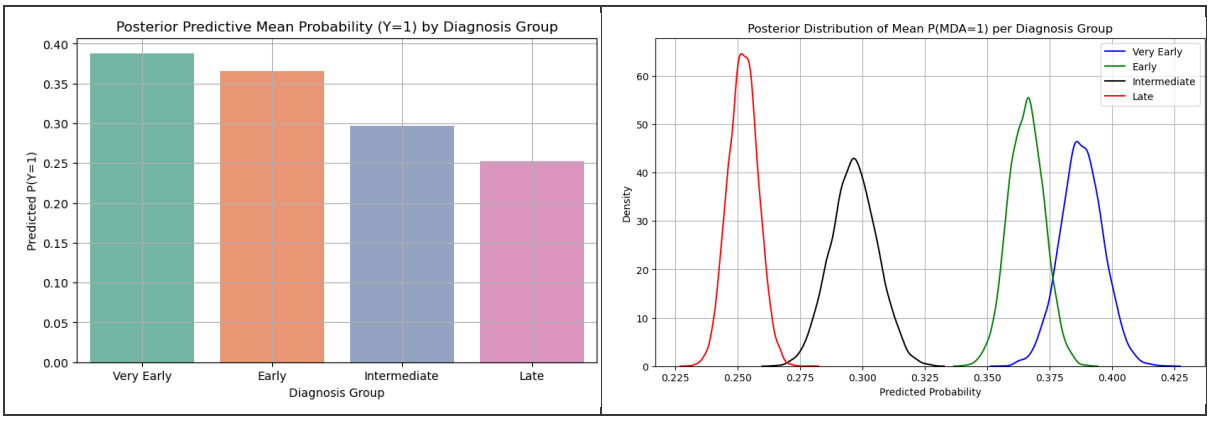


Figure 13 (Left) Posterior predictive mean probability of achieving Minimal Disease Activity (MDA) by diagnosis group. (Right) Posterior distributions of the group-level mean probability of MDA. Bars represent the group-level average predicted probability. Each curve reflects the uncertainty in P(MDA=1) for a diagnosis group, sampled from the Bayesian posterior.

Table 17 Posterior estimates of the mean probability of achieving Minimal Disease Activity (MDA) by diagnosis group and 95% Highest Density Intervals (HDI)

Diagnosis Group	**Mean** $P(\text{MDA} = 1)$	**95% HDI**
Very Early	0.388	[0.371, 0.405]
Early	0.366	[0.352, 0.379]
Intermediate	0.297	[0.278, 0.315]
Late	0.252	[0.240, 0.263]

Early diagnosis is associated with a higher and more rapidly achieved probability of reaching MDA, while patients diagnosed later remain systematically less likely to attain this clinical target. The strength and consistency of these group-level effects across time support the

clinical relevance of early detection in psoriatic arthritis management and provide robust probabilistic evidence that diagnostic timing significantly shapes therapeutic outcome

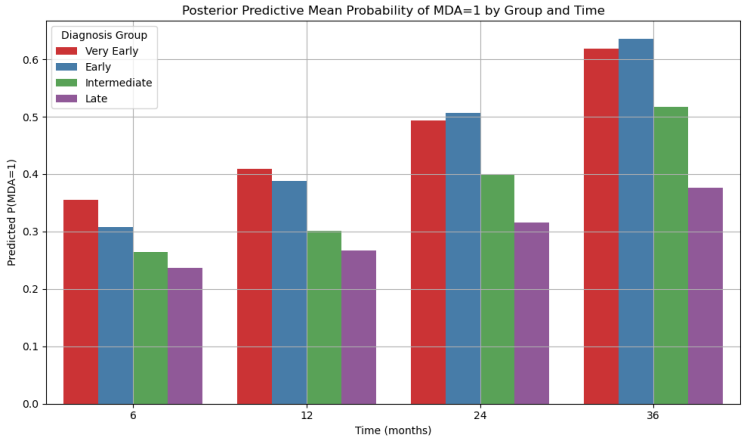


Figure 14 Posterior predictive mean probability of achieving Minimal Disease Activity (MDA) by diagnosis group and time point.

Reuma.pt cohort includes more difficult-to-treat cases which receive biologics treatment. This is reflected in the lower probabilities of achieving MDA we get from a Bayesian model on reuma.PT patients only (results in the **Appendix**) and is aligned with the previous results in **Figure 7**.

Highlight

Bayesian modeling of MDA outcomes revealed that patients diagnosed earlier consistently had a higher probability of achieving MDA across all timepoints. The diagnostic timing effect persisted and widened over patient observation time.

2.3.3.3 Population comparisons at Clinical Visits, from Baseline to 36-month visit

We analyzed longitudinal data from the DEPAR cohort to assess whether timing of diagnosis influences disease activity progression in patients with psoriatic arthritis (**Table 18**). Patients were stratified into two groups based on time from symptom onset to diagnosis: an Early group (if τ_d<1 year) and a Late group (if τ_d>1 year).

At baseline, there were no statistically significant differences between the Early and Late diagnosis groups across any of the assessed scores, including PASDAS, DAPSA, GRACE, and the SF-36 and its subdomains. This finding supports the comparability of groups at study entry and rules out baseline confounding due to pre-existing differences in disease burden.

However, from the 12-month visit (T12) onward, statistically significant differences began to emerge in several disease activity scores and functional domains. Patients in the Late diagnosis group consistently exhibited worse outcomes compared to those diagnosed early. These disparities persisted or widened over time, particularly in SF-36 domains reflecting physical and social function.

A notable example was observed in SF-36 social functioning at 36 months, where the Early group had a mean score of 83.59, compared to 75.54 in the Late group ($p < 0.05$). Similarly, GRACE and PASDAS scores showed lower disease activity in the Early group at later time points, reinforcing the cumulative impact of diagnostic delay.

These results suggest that delayed diagnosis is associated with a more adverse disease trajectory. The fact that significant differences only emerged after several months of follow-up implies that disease progression in the Late group cannot be attributed to worse initial disease

status, but rather reflects ongoing deterioration likely due to delayed treatment initiation and disease management. Our findings highlight the clinical value of timely diagnosis and intervention in altering long-term outcomes in psoriatic arthritis.

Table 18 Group-mean disease activity and quality of life scores over time in Early (<1 year) vs Late (>1 year) diagnosis groups in the DEPAR cohort.

Measure/ Visit	First Visit	T3	T6	T9	T12	T18	T24	T36
PASDAS	Early 4.24, Late 4.16	Early 3.34, Late 3.41	Early 2.99, Late 3.27	Early 2.81, Late 3.01	Early 2.64, Late 2.94	Early 2.70, Late 3.01	Early 2.60, Late 2.86	Early 2.53, Late 3.01
DAPSA	Early 19.65, Late 18.33	Early 11.95, Late 13.66	Early 10.66, Late 12.70	Early 9.71, Late 10.97	Early 8.87, Late 11.34	Early 9.23, Late 12.16	Early 9.10, Late 10.06	Early 7.75, Late 11.36
GRACE	Early 3.28, Late 3.32	Early 2.54, Late 2.73	Early 2.26, Late 2.61	Early 2.01, Late 2.35	Early 1.89, Late 2.36	Early 1.95, Late 2.40	Early 1.80, Late 2.15	Early 1.72, Late 2.18
sf36 vitality	Early 53.96, Late 50.60	Early 56.25, Late 51.70	Early 56.59, Late 52.75	Early 57.53, Late 54.30	Early 59.53, Late 53.46	Early 58.72, Late 53.31	Early 59.39, Late 53.58	Early 59.99, Late 53.74
sf36 mental	Early 71.99, Late 70.35	Early 72.48, Late 71.70	Early 72.98, Late 72.61	Early 74.16, Late 71.72	Early 76.04, Late 73.33	Early 75.11, Late 72.23	Early 76.50, Late 73.24	Early 78.06, Late 73.38
sf36 general health	Early 55.30, Late 50.69	Early 56.66, Late 51.68	Early 55.49, Late 52.38	Early 55.84, Late 52.81	Early 58.16, Late 52.59	Early 55.82, Late 52.98	Early 57.06, Late 54.01	Early 58.26, Late 54.63
sf36 role physical	Early 53.10, Late 53.88	Early 58.92, Late 57.14	Early 62.25, Late 58.84	Early 65.06, Late 60.51	Early 67.42, Late 60.42	Early 65.62, Late 60.75	Early 67.53, Late 61.40	Early 69.81, Late 62.03
sf36 bodily pain	Early 45.50, Late 47.53	Early 56.96, Late 54.55	Early 59.17, Late 57.42	Early 63.32, Late 58.60	Early 63.57, Late 58.71	Early 63.82, Late 58.79	Early 65.96, Late 59.30	Early 66.65, Late 58.53
sf36 role emotional	Early 72.31, Late 72.10	Early 76.18, Late 73.85	Early 77.93, Late 73.91	Early 78.26, Late 72.35	Early 79.50, Late 73.93	Early 79.78, Late 73.02	Early 78.80, Late 76.43	Early 80.57, Late 73.18
sf36 social functioning	Early 72.27, Late 70.86	Early 76.12, Late 70.81	Early 77.32, Late 72.22	Early 79.19, Late 74.15	Early 81.13, Late 75.87	Early 78.68, Late 73.66	Early 81.27, Late 75.39	Early 83.12, Late 75.44
sf36 physical functioning	Early 63.02, Late 62.04	Early 69.88, Late 66.77	Early 71.91, Late 67.88	Early 73.75, Late 70.75	Early 75.97, Late 70.39	Early 75.00, Late 70.22	Early 75.69, Late 71.74	Early 76.46, Late 67.75

Values represent group means at each visit timepoint (First Visit, T3, T6, T9, T12, T18, T24, T36) for various composite disease activity indices (PASDAS, DAPSA, GRACE) and SF-36 health-related quality of life domains. Early diagnosis group scores are shown in green when the difference with the

Late group is statistically significant ($p < 0.05$), based Mann-Whitney U tests using group-level summary statistics. Normality could not be assumed, according to Lilliefors test.

Highlight

Group-mean disease activity and quality of life scores were compared over time in Early (<1 year) vs Late (>1 year) diagnosis groups in the DEPAR cohort. Statistically significant differences begin to appear at T12 and persist or widen through T36, indicating more favorable disease trajectories in the Early diagnosis group.

2.3.3.4 Differences in disease activity in Late vs Early Diagnosis groups: PASDAS

Figure 15 presents the cumulative distribution functions (CDFs) of posterior predictive disease scores stratified by diagnosis group. Each curve represents the empirical CDF of predicted values from the Bayesian model, based on samples from the posterior predictive distribution of the outcome variable Y. The figure visualizes group-level differences in predicted disease burden at a population level.

The CDF for the "Late" diagnosis group is distinctly right-shifted compared to the others, indicating higher predicted PASDAS scores and thus worse disease control. This rightward shift reflects a stochastically larger distribution — at any given score threshold, a lower proportion of "Late" patients have improved disease activity than their earlier-diagnosed counterparts. The threshold line at 1.07 standardized units marks a clinical cutoff derived from prior literature (e.g., (Michielsens et al., 2022)), beyond which disease progression is considered clinically significant.

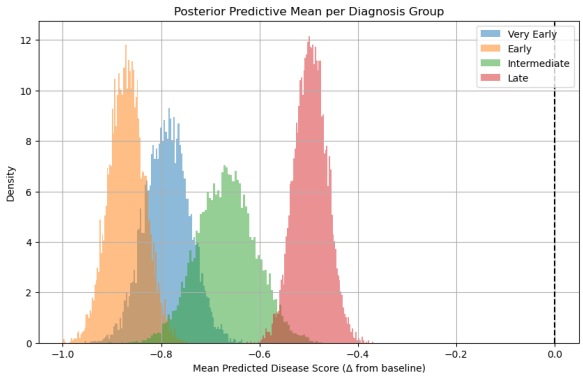


Figure 15 Cumulative Distribution Function for ΔPASDAS<x)

Figure 16 displays the posterior predictive distributions of the mean predicted change in PASDAS score (ΔPASDAS) from baseline for each diagnosis group. Each density represents

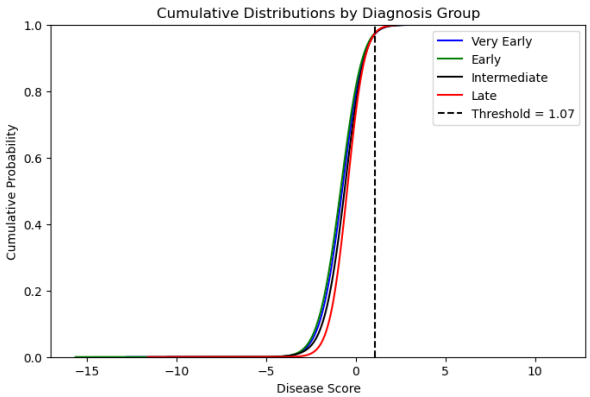


Figure 16 Posterior predictive distributions of the mean predicted change in PASDAS score (ΔPASDAS).

the distribution of expected group-level means sampled from the model's posterior predictive distribution $P\left(\overline{Y_g^{pred}} \mid \text{data}\right)$, where $\overline{Y_g^{\text{pred}}}$ is the group-specific average predicted change from baseline.

The distributions show a clear left-to-right gradient, with "Very Early" and "Early" groups having the most negative (improved) ΔPASDAS scores, and the "Late" group showing the least improvement. This ordering supports the hypothesis that earlier diagnosis is associated with greater improvement in disease activity over time. While some overlap exists (especially between Very Early and Early), the separation between Intermediate and Late groups is substantial. The vertical dashed line at 0 serves as a reference point for no change; all groups are left-shifted, indicating overall improvement, but to varying degrees.

These distributions reinforce that the probability of clinically meaningful improvement is higher and more certain in earlier diagnosis groups.

To further explore the temporal dynamics of disease progression, we evaluated cumulative distribution functions (CDFs) of predicted PASDAS scores at 6, 12, 24, and 36 months stratified by diagnosis timing (**Figure 17**). These plots illustrate a progressive divergence of the distributions across groups over time. Notably, by month 36, the separation between the "Late" and earlier diagnosis groups is more pronounced than at earlier timepoints, indicating a cumulative disadvantage in disease control for patients diagnosed later. This widening gap suggests that early diagnosis not only confers initial benefits but also contributes to a sustained trajectory of improved disease outcomes over time. These findings support the hypothesis that early intervention may alter the long-term disease course in psoriatic arthritis.

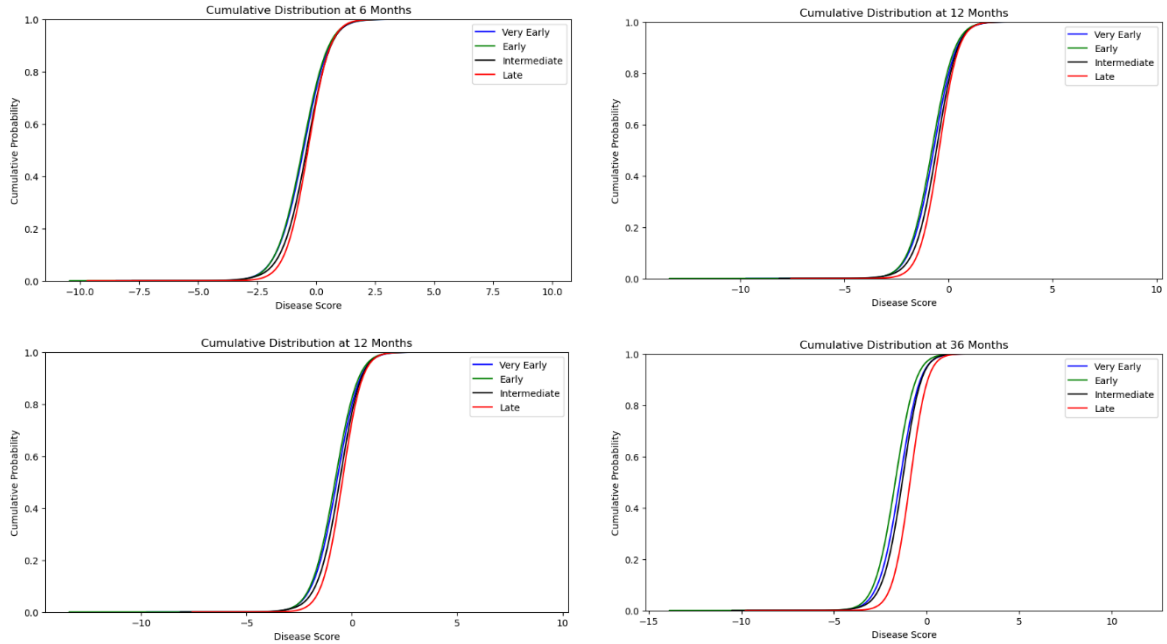


Figure 17 Cumulative distribution functions (CDFs) of posterior predictive PASDAS score changes from baseline, stratified by diagnosis group, at 6, 12, 24, and 36 months

Highlight:

From the Bayesian PASDAS analysis on DEPAR cohort, we observe the consistent separation between diagnosis groups, showing that patients diagnosed late had worse disease progression and suggesting that diagnostic timing is a powerful determinant of long-term outcomes in psoriatic arthritis, reinforcing the critical importance of early detection.

2.3.3.5 Treatment Persistence- Exploring patterns in the data

To explore how the timing of diagnosis relates to subsequent treatment persistence on DEPAR, we visualized patient transitions through EULAR treatment escalation phases across different diagnosis timing groups (**Figure 18**). Each stream in the Sankey diagram represents the flow of patients from their diagnosis timing category to the treatment phase they reached over the study period.

From the figure, no dominant or consistent trajectory pattern emerges that would suggest a straightforward relationship between the timing of diagnosis and the extent of treatment escalation. Patients across all diagnosis timing categories—whether diagnosed very early or late—appear distributed across the treatment phases, including the more advanced ones (Phases 4 and 5). While some early-diagnosed patients indeed remain in lower treatment phases, a substantial portion escalate rapidly. Conversely, some late-diagnosed patients remain in lower treatment phases, challenging the assumption that later diagnosis uniformly leads to more intensive treatment.

This heterogeneity suggests that other underlying factors may influence treatment persistence beyond diagnosis timing alone. These observations motivate a deeper investigation into latent patterns within patient trajectories. In the next section, we apply clustering techniques to uncover potential subgroups of patients with shared treatment progression patterns. Finally, we try a survival analysis model to investigate statistically significant effect of timing diagnosis on time-to-reach treatment phase.

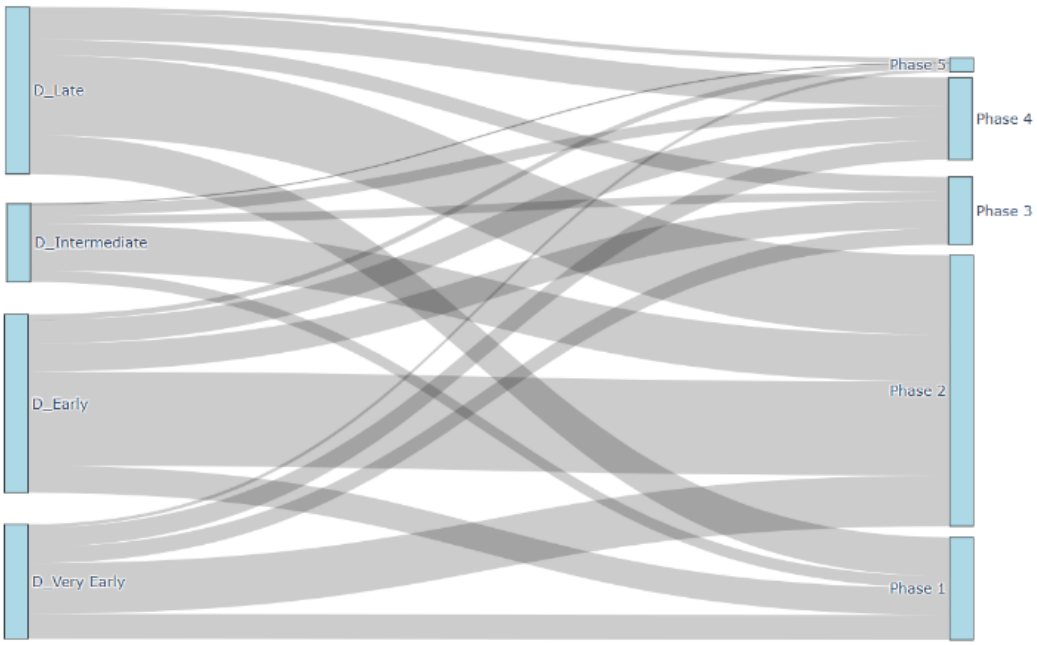


Figure 18 Sankey diagram of patient trajectories from diagnosis timing groups to EULAR treatment escalation phases.

2.3.3.6 Treatment Persistence- Clustering

We applied k-means clustering with k = 12 to segment DEPAR patients based on treatment trajectories and time-to-event patterns (e.g., initiation and escalation of drug therapies). The rationale behind selecting k = 12 was to capture heterogeneity in both diagnosis timing and

treatment outcomes, allowing clusters to reflect meaningful subgroups in the population — including stable responders, early dropouts, and late escalators.

From the stacked bar plots (**Figure 19**) and pie charts (**Figure 20**), we observe that:

- The clusters naturally align with treatment outcomes, particularly the most advanced phase reached.
- Diagnosis timing is broadly distributed across clusters, though certain patterns emerge.

There is some evidence that late-diagnosed patients (Late and Intermediate) are more likely to escalate to advanced disease phases, particularly Phase 4 or 5. Cluster 2 provides the clearest example of this trend, with 49 late-diagnosed patients contributing to a total of 147 individuals reaching Phase 4. However, this pattern is not consistent across all clusters. These findings indicate that while late diagnosis may increase the risk of disease progression, other factors, such as clinical phenotype, also play a critical role in shaping patient outcomes.

The clustering analysis reveals distinct patient trajectories, each representing a unique disease and treatment pathway. "Silent Watchers" (Clusters 0 and 11) are patients who were often diagnosed late but never initiated treatment, remaining in Phase 1 throughout—suggesting under-management or disengagement from care. "Early Stable Responders" (Clusters 1, 4, and 8) were diagnosed early, received timely intervention, and stabilized in Phase 2, reflecting effective disease control. "Slow Escalators" (Cluster 6) show gradual progression, possibly due to variable treatment response or fluctuating disease activity, eventually stabilizing after Phase 3. "Late-Diagnosis Escalators" (Cluster 2) were diagnosed at more advanced stages and required intensive management, with many reaching Phase 4. Finally, "Difficult-to-Treat" patients (Cluster 5) make up the only group dominated by Phase 5, yet their escalation is not strongly linked to late diagnosis—suggesting that disease phenotype or intrinsic treatment resistance may play a more significant role.

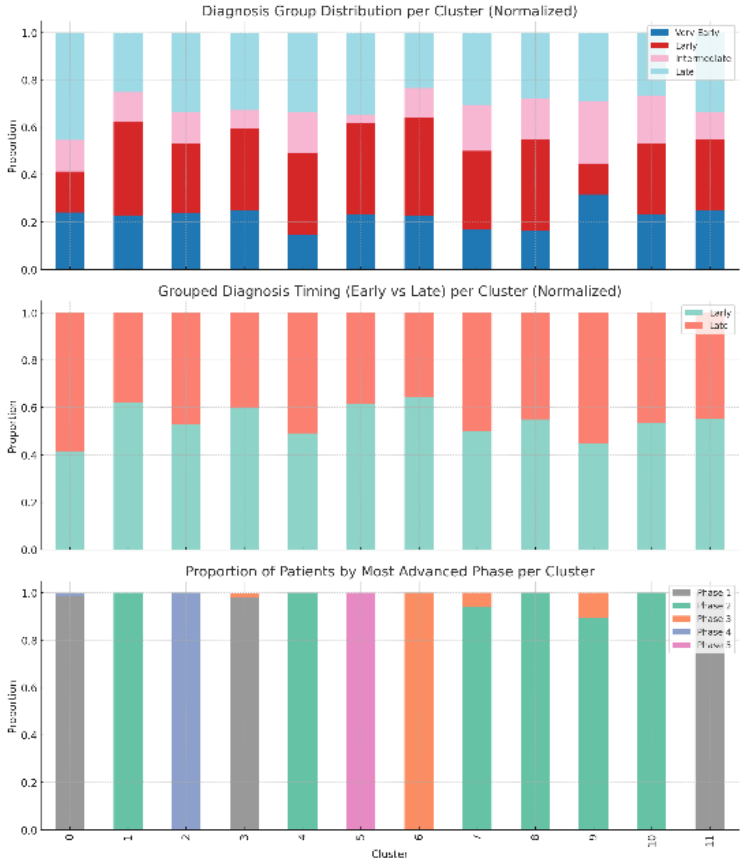


Figure 19 Diagnosis Timing and Disease Phase distribution across clusters.

(Top) Diagnosis groups per cluster. (Mid) Aggregated earl (Very Early, Early) and Late (Intermediate, Late) diagnosis per cluster. (Bottom) Proportion of patients assigned to their most advanced disease phase (Phases 1–5) per cluster

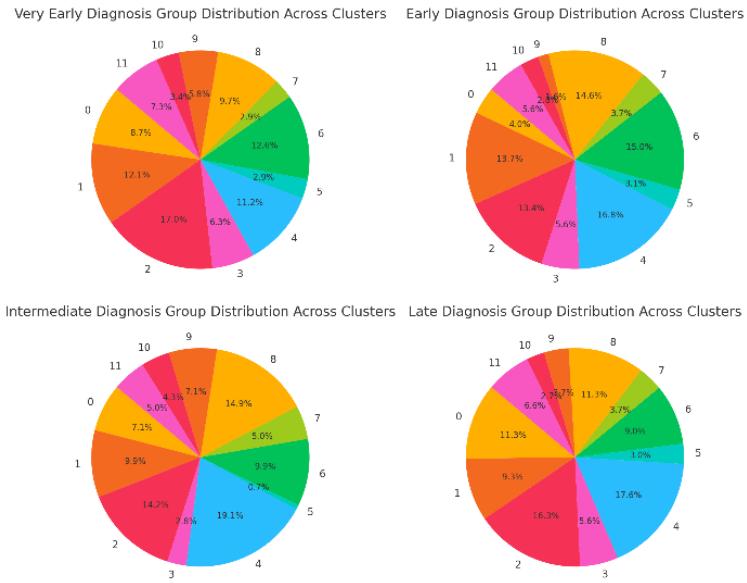


Figure 20 Distribution of Diagnosis Timing Groups Across Clusters

Highlight:
Within each diagnosis group — from Very Early to Late — patients follow diverse treatment trajectories. Some present with mild disease and stabilize quickly, while others escalate to more intensive phases and become difficult to treat. This variation suggests that diagnosis timing alone does not determine outcome — instead, other factors such as phenotype, comorbidities, or clinical presentation likely shape each patient's trajectory.

2.3.3.7 Treatment Persistence- Survival Analysis

Cox proportional hazards models on DEPAR cohort predicting the time-to-event for reaching EULAR treatment Phase 2, Phase 3, or Phase 4, are summarized in **Table 19**, which presents hazard ratios (HRs) with 95% confidence intervals. Statistically significant covariates ($p < 0.05$) are highlighted in green.

HRs for diagnosis timing are calculated relative to the 'Very Early' diagnosis group, which serves as the reference category. The effect of diagnosis timing is inconsistent across configurations, this further supports the mixed trajectory patterns observed in the Sankey diagram.

Concerning the kNN engineered feature, it captures the average trajectory of a patient's nearest clinical neighbors (based on diagnosis timing and covariates) and aims to summarize personalized, data-driven progression patterns. The hypothesis was that patients with similar clinical profiles might share common escalation risks not captured by standard covariates.

Across all Cox models and treatment phase thresholds, the kNN feature is not statistically significant. This suggests that the kNN-based clinical similarity measure, as constructed here, does not add predictive value beyond the explicitly modeled covariates. This could possibly mean that patient heterogeneity is not well captured by the available covariates.

Table 19 Cox proportional hazards models for the event of reaching EULAR treatment Phases 2, 3, or 4. The data in the table are in the format: Hazard ration (95% Confidence Interval).

Configuration	Age	BMI	Gender	Smoking	Early	Intermediate	Late	kNN feature	Enthesitis	Axial	Dactylitis	Oligo	Poly
Phase 2, kNN NO, Groups 2	1.012 (0.879, 1.165)	0.904 (0.658, 1.241)	1.041 (0.901, 1.203)	1.029 (0.879, 1.205)			0.970 (0.840, 1.120)		1.010 (0.831, 1.228)	1.021 (0.855, 1.220)	1.184 (0.987, 1.421)	1.623 (1.327, 1.983)	1.815 (1.525, 2.160)
Phase 2, kNN NO, Groups 4	1.017 (0.884, 1.170)	0.912 (0.662, 1.257)	1.042 (0.901, 1.205)	1.014 (0.863, 1.192)	1.233 (1.044, 1.455)	1.165 (1.004, 1.351)	1.028 (0.844, 1.252)		1.028 (0.847, 1.249)	1.016 (0.843, 1.225)	1.196 (1.000, 1.431)	1.610 (1.319, 1.966)	1.820 (1.536, 2.156)
Phase 2, kNN YES, Groups 2	1.033 (0.915, 1.166)	0.971 (0.777, 1.214)	1.033 (0.897, 1.191)	1.015 (0.880, 1.170)			1.002 (0.885, 1.133)	0.973 (0.827, 1.144)	0.976 (0.823, 1.158)	0.966 (0.707, 1.321)	1.162 (0.989, 1.364)	1.656 (1.378, 1.991)	1.842 (1.512, 2.243)
Phase 2, kNN YES, Groups 4	1.046 (0.928, 1.179)	0.967 (0.773, 1.210)	1.030 (0.905, 1.172)	1.002 (0.870, 1.154)	1.262 (1.085, 1.469)	1.193 (1.039, 1.371)	1.088 (0.917, 1.291)	0.926 (0.803, 1.068)	0.976 (0.823, 1.158)	0.965 (0.700, 1.329)	1.175 (1.005, 1.374)	1.659 (1.384, 1.990)	1.897 (1.560, 2.306)
Phase 3, kNN NO, Groups 2	0.693 (0.594, 0.808)	1.115 (0.749, 1.659)	0.855 (0.735, 0.994)	1.173 (0.963, 1.429)			0.961 (0.823, 1.121)		0.911 (0.699, 1.187)	0.994 (0.383, 2.580)	1.042 (0.825, 1.317)	1.613 (1.258, 2.068)	1.534 (1.205, 1.952)
Phase 3, kNN NO, Groups 4	0.693 (0.593, 0.810)	1.116 (0.747, 1.666)	0.853 (0.733, 0.993)	1.170 (0.957, 1.431)	0.966 (0.794, 1.176)	0.980 (0.822, 1.169)	0.930 (0.764, 1.132)		0.914 (0.701, 1.191)	0.998 (0.383, 2.597)	1.043 (0.825, 1.320)	1.620 (1.261, 2.082)	1.540 (1.207, 1.964)
Phase 3, kNN YES, Groups 2	0.696 (0.602, 0.805)	1.078 (0.898, 1.294)	0.905 (0.788, 1.038)	1.187 (1.010, 1.395)			0.937 (0.802, 1.093)	0.998 (0.780, 1.278)	0.992 (0.789, 1.249)	1.075 (0.817, 1.416)	1.068 (0.864, 1.320)	1.634 (1.290, 2.069)	1.648 (1.259, 2.157)
Phase 3, kNN YES, Groups 4	0.698 (0.603, 0.809)	1.081 (0.901, 1.296)	0.900 (0.784, 1.033)	1.180 (0.999, 1.393)	0.901 (0.760, 1.068)	0.923 (0.786, 1.085)	0.883 (0.734, 1.063)	1.013 (0.805, 1.276)	0.996 (0.790, 1.256)	1.082 (0.821, 1.425)	1.071 (0.867, 1.321)	1.651 (1.300, 2.098)	1.644 (1.262, 2.144)

Funded by the European Union

Configuration	Age	BMI	Gender	Smoking	Early	Intermediate	Late	kNN feature	Enthesitis	Axial	Dactylitis	Oligo	Poly
Phase 4, kNN NO, Groups 2	0.668 (0.543, 0.821)	1.113 (0.803, 1.541)	0.834 (0.675, 1.031)	1.144 (0.853, 1.533)			1.038 (0.845, 1.275)		0.929 (0.723, 1.193)	1.078 (0.750, 1.549)	0.923 (0.722, 1.180)	1.007 (0.770, 1.317)	1.140 (0.888, 1.465)
Phase 4, kNN NO, Groups 4	0.667 (0.542, 0.821)	1.115 (0.804, 1.546)	0.827 (0.667, 1.025)	1.140 (0.849, 1.531)	0.878 (0.676, 1.140)	0.953 (0.746, 1.218)	0.963 (0.744, 1.248)		0.934 (0.725, 1.202)	1.087 (0.754, 1.567)	0.919 (0.717, 1.177)	1.021 (0.776, 1.344)	1.159 (0.893, 1.503)
Phase 4, kNN YES, Groups 2	0.712 (0.593, 0.855)	1.146 (1.047, 1.254)	0.799 (0.659, 0.968)	1.149 (0.892, 1.479)			1.082 (0.908, 1.289)	0.819 (0.648, 1.036)	0.987 (0.788, 1.237)	1.091 (0.873, 1.363)	0.860 (0.669, 1.105)	1.097 (0.850, 1.417)	1.286 (1.006, 1.644)
Phase 4, kNN YES, Groups 4	0.695 (0.580, 0.832)	1.135 (1.024, 1.259)	0.836 (0.700, 0.998)	1.148 (0.895, 1.472)	0.874 (0.698, 1.095)	0.954 (0.782, 1.165)	0.980 (0.789, 1.218)	1.010 (0.802, 1.273)	0.981 (0.782, 1.232)	1.108 (0.887, 1.384)	0.926 (0.721, 1.191)	1.062 (0.823, 1.370)	1.202 (0.947, 1.526)

Green cells indicate statistical significance ($p < 0.05$).

3 Prospective data analysis: PDPID cohort

3.1 Objectives

3.1.1 Objectives preliminary analysis

(1) To describe the baseline characteristics of a sample of patients with PsA included in the PDPID study and to compare these characteristics between the participating countries (UK, Greece, Netherlands).

(2) To assess factors associated with disease activity in patients with PsA. These factors include PROs and environmental factors.

3.1.2 PDPID study objectives

The objectives of the PDPID study as listed in the protocol, include analysis tasks for T3.1, T3.2 and T3.5. For PDPID data analysis, T3.1 is the task involving analysis of clinical data.

3.1.2.1 Primary objectives

(1) To develop and internally validate a novel and interpretable machine learning model for detecting flare in PsA patients using integrated accelerometer data, keystroke dynamics and screen time metrics (i.e., digital biomarker) to assess changes in their physical activity patterns against clinical defined flare by the rheumatologist. Accelerometer data is captured by both smartphone and smartwatch.

(2) To develop and internally validate machine learning models that capitalize on sleep, fatigue, pain, stress, mechanical stress, composition of gut microbiome, genetic risk and environmental exposure for flare prediction (either clinically established or evaluated by the digital biomarker) in patients with PsA.

3.1.2.2 Secondary objectives

(1) To assess construct validity of the novel and interpretable machine learning model for detecting flare in PsA patients using integrated accelerometer data, keystroke dynamics and screen time metrics (digital biomarker) to assess changes in physical activity patterns against the continuous measure of clinical composite scores of disease activity and impact of disease used by the rheumatologist and impact of disease as reported by the patient

(2) To develop and internally validate a novel and interpretable machine learning model for changes in joint and skin appearance that relates to flare in PsA patients using video analysis of hand, posture and gesture and photos of the hands and feet against clinically defined flare by the rheumatologist

(3) To determine intrapersonal reliability of the AI-driven digital biomarker system

(4) To determine clinically relevant changes in the AI-driven digital biomarker system

(5) To determine minimal detectable difference in the AI-driven digital biomarker system

(6) To assess the intrapersonal variation of stress, mechanical stress and changes in gut microbiome on the occurrence of flare

(7) To assess genetic contribution to disease activity and pain

(8) To evaluate costs and effects of the digital biomarker of future care to current care.

Funded by the
European Union

(9) To evaluate the compliance and satisfaction of the users with the smartphone- and smartwatch-based measurement of disease activity and flare

3.2 Methods

This section describes the methods used for the preliminary analysis of data from the PDPID study.

3.2.1 Study design and sample

We conducted a cross-sectional study using baseline data from a sample of patients with PsA included in the PDPID cohort from Netherlands, UK and Greece. Patients were included from multiple centers in UK, Netherlands, and Greece. For the preliminary analysis, data from Portugal has not been yet harmonized with the data from the other three countries (will be included in future analyses).

The PDPID study was approved by the Medical Research Ethics Committee (MREC) of each country and by the ethics committees of the participating centers. Patients provided a written informed consent prior to inclusion.

The inclusion criteria include: (i) 18 years or older and competent (ii) diagnosed with PsA (iii) using a smart phone (iv) agree to use smartwatch and (v) good command of local language. The exclusion criteria include: (i) less than 18 years of age (ii) incapacitated patients.

3.2.2 Primary study outcomes

The primary study outcome is Absence of flare in PsA evaluated every three months. For the preliminary analysis, the questionnaires are assessed at baseline.

3.2.2.1 Patients

‘At this time, is your psoriatic arthritis in remission, if this means: you feel your disease is as good as gone?’

‘At this time, are you in low disease activity, if this means: your disease is in low activity but it’s not as good as gone?’

3.2.2.2 Doctors

At this time, is the psoriatic arthritis in remission, if this means: the absence of clinical and laboratory evidence of significant inflammatory disease activity?’

‘At this time, is the psoriatic arthritis in low or minimal disease activity?’

3.2.2.3 Patient Acceptable Symptom State (PASS)

‘If you were to remain for the next few months as you were during the last 48 hours, would this be acceptable or unacceptable for you?’ yes/ no

3.2.3 Secondary study outcomes

Based on clinical evaluation and the patient reported outcomes the following composites of disease activity and impact of disease are calculated at baseline and during each follow-up period every three months. For the preliminary analysis, the scores are assessed at baseline.

3.2.3.1 Disease activity composite scores

MDA is defined by patient meeting 5 out of 7 of the following criteria: (i) TJC ≤1; (ii) SJC ≤1; (iii) PASI ≤ 1 or BSA ≤3%; (iv) tender entheseal points ≤ 1 (v) patient pain visual analogue scale VAS ≤15; (vi) patient global activity VAS ≤20; (vii) HAQ ≤0.5.

PASDAS calculated as a composite score using the following criteria: physician global VAS, patient global VAS, SF36 PCS, SJC, TJC, Leeds Enthesitis Index, tender dactylitis count, and CRP.

DAPSA calculated as a summation score of the following: tender and swollen joints (TJC68, SJC66), patient global assessment VAS, patient pain VAS and CRP.

3.2.3.2 Impact of disease

Likert scale questions on a daily basis from the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire and a 10-point Likert scale for severity of stiffness.

For the preliminary analysis, PsAID pain, fatigue, sleep and morning stiffness are assessed as part of baseline PROs.

3.2.3.3 Flare questionnaire

Flare questionnaire is asked at baseline, follow-up visits, and when patients experience disease flare in which they seek help from the rheumatologist. For the preliminary analysis, flare questionnaire is assessed at baseline.

3.2.3.3.1 Doctors

'At this time, is the disease in flare (i.e., significantly worsened/more active compared to usual)?' yes/no

3.2.3.3.2 Patients

'At this time, are you having a flare of your psoriatic arthritis, if this means the symptoms are worse than usual?' yes/no

3.2.4 Other study parameters

Only the study parameters relevant to the preliminary analysis are included in this section.

3.2.4.1 Stress

For the preliminary analysis, stress is assessed as part of baseline PRO using the perceived stress scale (PSS) questionnaire. The PSS score ranges from 0 to 40 with higher scores indicating higher perceived stress.

3.2.4.2 Environmental factors

For the preliminary analysis, we assessed the average daily weather and air pollution exposure at the patient home location on the day of the baseline visit. Weather variables included temperature and humidity. Air pollution variables included NO_2, O_3, PM_{10}, PM_{25}, and SO_2.

3.2.5 Data analysis

Descriptive statistics were used to summarize the characteristics of the study sample. Median with interquartile ranges is used for continuous variables, and N with percentage for categorical variables. Comparisons between countries are performed using Kruskal-Wallis test for continuous variables and chi-square test for categorical variables.

Pairwise correlations were used to correlate the disease activity measures (flare, remission, LDA, PASS). We assessed the impact of flare by analysing the associations between flare and PROs (pain, fatigue, sleep and stiffness) using linear regression models. To analyse associations between stress and environmental factors and disease activity, linear and logistic regression models were used. Since PsAID questions are assessed as aspects of disease impact they were not included as independent variables in the regression models for disease activity.

3.3 Results

This section presents the preliminary results from the PDPID study.

3.3.1 Baseline characteristics

Baseline characteristics of the study population and comparison between countries are shown in **Table 20**. A total of 124 patients with PsA were included. The median age of patients was 55 years. Around half of the study population was females. The median education level was in the moderate-to-high range (13 years). Seventeen percentage of patients were in flare according to the physician and 10% of patients reported flare. The most common disease phenotype was polyarticular followed by oligoarticular. The median pain, fatigue and sleep levels were 2, while median stiffness was 3. There were significant differences between countries in number of patients in flare, LDA, PASS, PASDAS, MDA, disease phenotype, swollen and tender joint count, pain, sleep and stiffness.

Table 20 Baseline characteristics of study population

Baseline characteristics	Total N=124	Netherlands N=30	UK N=42	Greece N=52	P-value
Demographics					
Age in years, median (IQR)	55 (44, 61)	52 (45, 60)	58 (49, 62)	54 (41, 61)	0.247
Missing, n(%)	1(1%)	0	1(2%)	0	
Sex (F), n(%)	64 (51%)	14 (46%)	25 (50%)	25 (48%)	0.370
Missing, n(%)	1(1%)	0	1(2%)	0	
Education in years, median (IQR)	13 (12, 16)	13(12, 16)	13 (12, 16)	13 (12, 16)	0.471
Missing, n(%)	2 (2%)	1 (3%)	1(2%)	0	
Flare					
Flare (physician)	21(17%)	8(27%)	8(19%)	5(10%)	**0.006**
Missing, n(%)	*5(4%)*	*4(13%)*	*1(2%)*	*0*	
Flare (patient) – Yes, n(%)	12(10%)	-	7(17%)	5(10%)	**<0.0001**
Missing, n(%)	32(26%)	30(100%)	1(2%)	1(2%)	
Remission (physician) – Yes, n(%)	72(58%)	16(53%)	25(60%)	31(60%)	0.056
Missing, n(%)	5(4%)	4(13%)	1(2%)	0	
Remission (patient) – Yes, n(%)	47(38%)	4(13%)	12(29%)	31(60%)	**<0.0001**
Missing, n(%)	6(5%)	5(17%)	1(2%)	0	
LDA (physician) – Yes, n(%)	62(50%)	21(70%)	26(62%)	15(29%)	**<0.0001**
Missing, n(%)	5(4%)	4(13%)	1(2%)	0	
LDA (patient) – Yes, n(%)	64(52%)	17(57%)	32(76%)	15(29%)	**<0.0001**
Missing, n(%)	6(5%)	5(17%)	1(2%)	0	
PASS (patient) – Yes, n(%)	73 (59%)	17 (57%)	27 (64)	29 (56)	**0.037**
Missing, n(%)	7(6%)	5(7%)	1(2%)	1(2%)	
Disease activity scores					
PASDAS, median (IQR)	3 (2, 4)	3.8 (3, 4)	4 (3.7, 6)	3 (2, 4)	**0.001**
Missing, n(%)	57 (46%)	14 (47%)	34 (81%)	9 (17%)	
DAPSA, median (IQR)	43 (23, 101)	56 (38, 101)	86 (37, 166)	33 (12, 83)	0.09
Missing, n(%)	63 (51%)	21 (70%)	34 (81%)	8 (15%)	
MDA, n(%)	35 (28%)	5 (17%)	1 (2%)	29 (56%)	**<0.0001**
Missing, n(%)	51 (41%)	16 (53%)	30 (71%)	5 (10%)	

Baseline characteristics	Total N=124	Netherlands N=30	UK N=42	Greece N=52	P-value
Clinical assessments & PROs					
Disease phenotype, n(%)					
Polyarticular	50(40%)	-	27(64%)	23 (44%)	**<0.0001**
Oligoarticular	20(16%)	-	1(2%)	19 (37%)	
Monoarticular	5(4%)	-	5(12%)	0	
Entheseal	2(2%)	-	1(2%)	1 (2%)	
Dactylitis	3(2%)	-	3(7%)	0	
Axial	9(28%)	-	0	9 (17%)	
Missing, n(%)	35 (28%)	29 (100%)	5(12%)	0	
SJC, median (IQR)	0 (0, 1)	0 (0, 1)	0 (0, 2)	0 (0,0)	**0.048**
Missing, n(%)	0	0	0	0	
TJC, median (IQR)	2 (0, 7)	0 (0, 2)	4 (2, 10)	1 (0, 5)	**0.0002**
Missing, n(%)	0	0	0	0	
Pain, median (IQR)	2 (1, 5)	3 (1, 5)	6 (2, 7)	2 (1, 3)	**0.008**
Missing, n(%)	35 (28%)	1 (3%)	28 (67%)	6 (12%)	
Fatigue, median (IQR)	2 (1, 4)	1 (0, 3)	4 (1,5)	3 (1, 4)	0.052
Missing, n(%)	35 (28%)	1 (3%)	28 (67%)	6 (12%)	
Sleep disturbance, median (IQR)	2 (0, 4)	1 (0, 4)	4 (2, 8)	2 (0, 4)	**0.018**
Missing, n(%)	35 (28%)	1 (3%)	28 (67%)	6 (12%)	
Stiffness, median (IQR)	3 (1, 6)	4 (2, 7)	5 (3, 7)	2 (1, 4)	**0.0001**
Missing, n(%)	3 (2%)	2 (7%)	1(2%)	0	
PSS score, median (IQR)	14 (10, 19)	13 (8, 16)	15 (10, 20)	16 (10, 20)	0.094
Missing, n(%)	36 (29%)	2 (7%)	28 (67%)	6 (12%)	

Notes. Inclusion dates: Netherlands (Sep 2024- March 2025), UK (Oct 2024 – May 2025), Greece (Oct 2024 – April 2025). P-value is for the comparisons between the three countries.

Highlight

The median age of patients was 56 years, with half of patients being females. At baseline, 17% of patients with PsA experienced flare according to physician and 10% according to the patient. The most common disease phenotype was polyarticular. There were significant differences between countries in the majority of disease activity variables, clinical assessment and PROs.

3.3.2 PROs and PsA disease activity

Pairwise correlation between the flare measures (flare, remission, LDA, PASS) are shown in **Table 21**. Physician-reported flare was significantly positively correlated with patient-reported flare, and negatively with patient- and physician-reported remission and LDA, and PASS. Patient-reported flare was positively correlated with physician-reported flare, and negatively with patient- and physician-reported remission, patient-reported LDA, and PASS.

Table 21. Pairwise correlation coefficients between PsA flare measures

	Flare (phys)	Flare (pat)	Rem (phys)	Rem (pat)	LDA (phys)	LDA (Pat)	PASS (pat)
Flare (phys)	1						
Flare (pat)	**0.95*** **	1					

	Flare (phys)	Flare (pat)	Rem (phys)	Rem (pat)	LDA (phys)	LDA (Pat)	PASS (pat)
Rem (phys)	**-0.48*** **	**-0.42*** **	1				
Rem (pat)	**-0.37*** **	**-0.36*** **	**0.61*** **	1			
LDA (phys)	**-0.22***	**-0.27** **	**-0.19***	**-0.32** **	1		
LDA (Pat)	**-0.36** **	**-0.33** **	-0.07	**-0.40*** **	**0.68*** **	1	
PASS (pat)	**-0.49*** **	**-0.49*** **	**0.72*** **	**0.55*** **	-0.034	0.07	1

Pairwise correlation. Rem, remission; pat, patient; phys, physician; PASS, patient acceptable symptom state. *** p<0.00001. ** p<0.01. *p<0.05

Associations between flare and PROs are shown in **Table 22**. Physician- and patient-reported flare were both positively associated with higher levels of pain, fatigue, sleep disturbance, and stiffness (p-values <0.01).

Table 22. Associations between flare and PROs in PsA.

PRO	Flare (physician)	Flare (patient)
Pain (β, 95%CI)	**2.29 (1.06, 3.52)*** **	**2.87 (1.16, 4.59)** **
Fatigue (β, 95%CI)	**1.54 (0.26, 2.81)***	**2.83 (1.15, 4.51)** **
Sleep disturbance (β, 95%CI)	**1.64 (0.27, 3.04)***	**2.98 (1.18, 4.77)** **
Stiffness (β, 95%CI)	**2.88 (1.63, 4.12)*** **	**3.38 (1.78, 4.97)*** **

Linear regression. *** p<0.0001, ** p<0.01, p<0.05

Higher PSS score, i.e., higher perceived stress levels was significantly associated increased odds for patient-reported flare, higher PASDAS and DAPSA and lower odds to achieve MDA (p-values <0.05) (**Table 23**).

Table 23. Associations between stress and PsA disease activity

Disease activity measure	PSS score
Flare-physician (OR, 95%CI)	1.10 (0.99, 1.23)
Flare-patient (OR, 95%CI)	**1.21 (1.01, 1.44)***
PASDAS (β, 95%CI)	**0.07 (0.01, 0.14)***
DAPSA (β, 95%CI)	**4.12 (1.63, 6.61)** **
MDA (OR, 95%CI)	**0.89 (0.81, 0.98)***

Linear regression. PSS, perceived stress scale. ** p<0.01, * p<0.05

Highlight

Physician- and patient-reported were highly positively correlated. Physician and patient-reported flare were negatively correlated with remission, LDA, and PASS. Flare was associated with worse PROs (pain, fatigue, sleep disturbance, stiffness). Stress was associated with higher disease activity (patient-reported flare, PASDAS DAPSA, lower odds for achieving MDA).

3.3.3 Environmental conditions and PsA disease activity

Based on **Table 24** it can be observed that: (i) The average temperature and humidity on the day of the study visit were not significantly associated with disease activity; (ii) Air pollutant nitrogen dioxide (NO_2) was associated with higher PASDAS and DAPSA, and lower odds for

achieving MDA (p-values <0.05), whereas no significant associations were observed for physician and patient-reported flare; (iii) Ozone (O_3) was associated with lower PASDAS and DAPSA (p-values <0.05), whereas no significant associations were observed for physician and patient-reported flare, and MDA; (iv) PM_{10}, PM_{25}, and SO_2 air pollutants were not significantly associated with disease activity.

Table 24. Associations between environmental factors and PsA disease activity

	Temp	**Humidity**	**NO_2**	**O_3**	**PM_{10}**	**PM_{25}**	**SO_2**
Flare-physician (OR, 95%CI)	1.06 (0.99, 1.15)	0.96 (0.92, 1.01)	1.01 (0.96, 1.06)	1.00 (0.97, 1.04)	0.99 (0.92, 1.05)	0.97 (0.90, 1.06)	0.80 (0.48, 1.31)
Flare-patient (OR, 95%CI)	1.08 (0.98, 1.20)	0.98 (0.90, 1.06)	0.96 (0.84, 1.09)	1.01 (0.94, 1.08)	0.99 (0.88, 1.10)	0.98 (0.85, 1.11)	0.59 (0.18, 1.94)
PASDAS (β, 95%CI)	0.008 (-0.04, 0.05)	0.03 (-0.1, 0.06)	**0.03 (0.005, 0.06)***	**-0.03 (-0.04, -0.009)****	0.01 (-0.02, 0.04)	0.006 (-0.03, 0.04)	-0.05 (-0.20, 0.11)
DAPSA (β, 95%CI)	0.34 (-1.57, 2.24)	0.70 (-1.16, 2.56)	**1.29 (0.06, 2.52)***	**-0.98 (-1.76, -0.20)***	0.23 (-1.13, 1.59)	0.11 (-1.35, 1.57)	-0.58 (-6.68, 5.53)
MDA (OR, 95%CI)	1.00 (0.93, 1.06)	0.99 (0.93, 1.06)	**0.95 (0.91, 0.99)***	1.03 (0.99, 1.06)	1.00 (0.95, 1.05)	1.01 (0.96, 1.06)	1.05 (0.83, 1.32)

Linear regression. ** p<0.01, * p<0.05

Highlight

On the day of study visit, higher Nitrogen dioxide levels were associated with higher disease activity (PASDAS, DAPSA, lower odds for MDA), whereas higher ozone levels were associated with lower disease activity (PASDAS, DAPSA). Temperature, humidity, and air pollutants PM_{10}, PM_{25}, and SO_2 on the day of study visit were not associated with disease activity.

3.4 Data analysis timeline

This section describes the proposed data analysis plan for the PDPID study. Task T3.1 will also provide data analysis outcomes for Task 3.5.

3.4.1 Study population characteristics

We will analyse the baseline study sample characteristics using descriptive statistics. We will also analyse the flare rates at baseline and follow-up periods in the total study sample, and by country, and compare them. The rates will be summarized using descriptive statistics, and comparisons between countries will be performed using regression models for single time points and mixed-effects models for longitudinal data.

Timeline: Harmonized data will be updated monthly, and the aforementioned analysis will be conducted regularly to present the most recent results.

3.4.2 Clinical assessments and PROs

We will analyse the associations between age, sex, years of disease, perceived stress, BMI, abdominal circumference and disease activity (composite scores and flare questions). We will assess the impact of flare by analysing the associations between flare and PROs. Regression models will be used for single time points and mixed-effects models for longitudinal data.

Timeline: Harmonized data will be updated monthly, and the aforementioned analysis will be conducted regularly to present the most recent results.

3.4.3 Biological data

3.4.3.1 Hair cortisol data

Hair cortisol is collected at T3, and optional at the other time-points. We will analyse the association between hair cortisol levels and the PSS scale using regression models at T3.

We will analyse the associations between hair cortisol at T3 and disease activity (composite scores and flare questions) at T3 and subsequent time-points. Regression models will be used for single time points and mixed-effects models for longitudinal data.

Timeline: Hair cortisol at T3 will be analysed in batches by the lab. [Expected start date of analysis: depends on the availability of the T3 visit first batch; Jan 2026]

3.4.3.2 Genetic data

Salivary DNA is collected at baseline. We have identified a list of genetic polymorphisms to investigate in relation to disease activity and pain levels. Patients will be categorized based on the presence or absence of these polymorphisms. We will analyse the associations between the genetic polymorphisms and disease activity and pain levels. Regression models will be used for single time points and mixed-effects models for longitudinal data.

Timeline: DNA will be analysed in the lab after the end of the data collection period (Dec 2025) for all baseline visits. The laboratory analysis and data processing are expected to take approximately three months before the dataset is available [Expected start date of analysis: April 2026].

3.4.3.3 Gut microbiome data

For the total study population, stool samples are collected are collected at baseline, and when patients experience flare. 16S rRNA gene sequencing is performed by the lab to obtain gut microbiome data. Microbial diversity will be analysed using the alpha- and beta- diversity metrics. Differential abundance analysis will be performed to identify specific taxa associated with disease activity (e.g., flare vs no flare state; high vs low disease activity score; MDA vs non-MDA).

Timeline: Gut microbiome will be analysed in batches. The plan is to analyse at the end of data collection period (Dec 2025) for baseline inclusions, and July for the rest of samples. The laboratory analysis and data processing are expected to take approximately three months before the dataset is available. [Expected start date of analysis: April 2026]

3.4.4 Environmental data

Environmental data is collected from openweathermap.org. Analysis will be performed at the patient home and work location. For the weather data, we will analyse the associations between the average weather conditions on the day of the visit, and average weekly and monthly weather conditions prior to the visit, and disease activity. We will assess the association between weather conditions when patient experiences flare as compared to period without a flare, to determine whether changes in weather may trigger flare. For air pollution, we will assess the associations between short-term exposure (day of visit), and depending on data availability, long-term exposure (annual), to air pollutants and disease activity.

Timeline: Environmental data is extracted automatically via API. Harmonized data will be updated monthly, so this analysis will be conducted regularly to present the most recent results.

4 Conclusions

PsA involves interactions between genetic, immunologic, and environmental factors. Patients with PsA experience periods of disease exacerbation that can influence their quality of life. Knowledge on the triggers of inflammation (disease activity) is very limited. Several factors can impact PsA prognosis including time to diagnosis and treatment. Besides, factors such as sex differences impact the disease manifestations and treatment response, add to the complexity of managing the disease.

The key takeaways of the deliverable are:

DEPAR and Reuma.pt data (retrospective)

- Analysis of retrospective data revealed differences between DEPAR and Reuma.pt such as in disease activity measures, phenotype and duration.
- Differences between males and females were observed across multiple disease-related measures importantly, disease activity, joint count, symptom and disease duration, phenotype, pain levels, and treatment duration and retention.
- Early PsA diagnosis is associated with better disease activity (i.e., lower disease activity) and quality of life (e.g., certain SF-36 domains) outcomes.
- Our results suggest that diagnosis timing alone does not determine treatment trajectory of the patients.

PDPID data (prospective)

- At baseline, 17% of patients with PsA experienced flare according to physician and 10% according to the patient. There were significant differences between countries in the majority of variables related to disease activity, clinical assessment and PROs.
- The majority of flare measures (e.g., flare, remission, LDA) were significantly correlated.
- Flare is associated with worse pain, fatigue, sleep disturbance, stiffness levels.
- Higher perceived stress was associated with higher disease activity.
- Higher NO_2 exposure is associated with higher disease activity, while higher O_3 is associated with lower disease activity. Temperature and humidity were not associated with disease activity.

For the retrospective data analysis, the next steps include conducting a more detailed analysis and considering publication of these research findings. In addition, the current analysis did not include the MONITOR data set, which is intended to be included in future analyses.

For the PDPID study (prospective data analysis), a data analysis plan has been established. The analysis will be expanded to the four countries (Netherlands, UK, Greece, and Portugal) with more participants and follow-up data. Further analysis of the PDPID clinical data will be conducted to achieve the study objectives, inform task T3.5, and identify triggers of inflammation. This will also include analysis of the genetic and microbiome data. As part of this process, the quality of the data and any issues related to data missingness will be examined.

References

Baral, P., Udit, S., & Chiu, I. M. (2019). Pain and immunity: implications for host defence. *Nature Reviews Immunology*, *19*(7), 433-447. https://doi.org/10.1038/s41577-019-0147-2

Besedovsky, L., Lange, T., & Haack, M. (2019). The Sleep-Immune Crosstalk in Health and Disease. *Physiol Rev*, *99*(3), 1325-1380.

Cañete, J. D., Tasende, J. A. P., Laserna, F. J. R., Castro, S. G., & Queiro, R. (2020). The Impact of Comorbidity on Patient-Reported Outcomes in Psoriatic Arthritis: A Systematic Literature Review. *Rheumatol Ther*, *7*(2), 237-257.

Cubino, N., Montilla, C., Usategui-Martin, R., Cieza-Borrela, C., Carranco, T., Calero-Paniagua, I., Quesada, A., Canete, J. D., Queiro, R., Sanchez, M. D., Hidalgo, C., Martinez, O., Del Pino-Montes, J., Diaz-Alvarez, A., & Gonzalez-Sarmiento, R. (2016). Association of IL1Beta (-511 A/C) and IL6 (-174 G > C) polymorphisms with higher disease activity and clinical pattern of psoriatic arthritis. *Clin Rheumatol*, *35*(7), 1789-1794. https://doi.org/10.1007/s10067-016-3301-2

De Benedittis, G., Latini, A., Conigliaro, P., Triggianese, P., Bergamini, A., Novelli, L., Ciccacci, C., Chimenti, M. S., & Borgiani, P. (2022). A multilocus genetic study evidences the association of autoimmune-related genes with Psoriatic Arthritis in Italian patients. *Immunobiology*, *227*(4), 152232. https://doi.org/10.1016/j.imbio.2022.152232

Dhabhar, F. S. (2014). Effects of stress on immune function: the good, the bad, and the beautiful. *Immunologic Research*, *58*(2), 193-210. https://doi.org/10.1007/s12026-014-8517-0

Duruöz, M. T., Gezer, H. H., Nas, K., Kılıç, E., Sargın, B., Kasman, S. A., Alkan, H., Şahin, N., Cengiz, G., Cüzdan, N., Gezer İ, A., Keskin, D., Mülkoğlu, C., Reşorlu, H., Ataman, Ş., Bal, A., Küçükakkaş, O., Yurdakul, O. V., Melikoğlu, M. A., . . . Tuncer, T. (2021). Gender-related differences in disease activity and clinical features in patients with peripheral psoriatic arthritis: A multi-center study. *Joint Bone Spine*, *88*(4), 105177.

Fries, J. F., Spitz, P., Kraines, R. G., & Holman, H. R. (1980). Measurement of patient outcome in arthritis. *Arthritis Rheum*, *23*(2), 137-145.

Gossec, L., Kerschbaumer, A., Ferreira, R. J. O., Aletaha, D., Baraliakos, X., Bertheussen, H., Boehncke, W. H., Esbensen, B. A., McInnes, I. B., McGonagle, D., Winthrop, K. L., Balanescu, A., Balint, P. V., Burmester, G. R., Cañete, J. D., Claudepierre, P., Eder, L., Hetland, M. L., Iagnocco, A., . . . Smolen, J. S. (2024). EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2023 update. *Ann Rheum Dis*, *83*(6), 706-719.

Gottlieb, A. B., McInnes, I. B., Rahman, P., Kollmeier, A. P., Xu, X. L., Jiang, Y., Sheng, S., Shawi, M., Chakravarty, S. D., Lavie, F., & Mease, P. J. (2023). Low rates of radiographic progression associated with clinical efficacy following up to 2 years of treatment with guselkumab: results from a phase 3, randomised, double-blind, placebo-controlled study of biologic-naïve patients with active psoriatic arthritis. *RMD Open*, *9*(1).

Gratacós, J., Galíndez, E., & Otón, T. (2021). Is obesity a predictor for lack of response to treatment in psoriatic arthritis? A systematic review

¿Es la obesidad un factor predictivo de falta de respuesta al tratamiento en la artritis psoriásica? Actualización de una revisión sistemática. *Reumatol Clin (Engl Ed)*, *17*(5), 268-278.

Healy, P. J., & Helliwell, P. S. (2008). Measuring clinical enthesitis in psoriatic arthritis: assessment of existing measures and development of an instrument specific to psoriatic arthritis. *Arthritis Rheum*, *59*(5), 686-691.

Husni, M. E., Merola, J. F., & Davin, S. (2017). The psychosocial burden of psoriatic arthritis. *Seminars in Arthritis and Rheumatism*, *47*(3), 351-360. https://doi.org/https://doi.org/10.1016/j.semarthrit.2017.05.010

Kasiem, F. R., Luime, J. J., Vis, M., Kok, M. R., Wervers, K., Gerards, A. H., Appels, C., van der Graaff, W. L., Starmans-Kool, M., Goekoop-Ruiterman, Y., van Groenendael, J., Korswagen, L. A., Veris-van Dieren, J. J., Hazes, J., & Tchetverikov, I. (2021). Lessons learned from clinical phenotypes in early psoriatic arthritis: the real-world Dutch south west Early Psoriatic ARthritis study. *Scand J Rheumatol*, *50*(2), 124-131. https://doi.org/10.1080/03009742.2020.1803398

Lembke, S., Macfarlane, G. J., & Jones, G. T. (2024). The worldwide prevalence of psoriatic arthritis—a systematic review and meta-analysis. *Rheumatology*. https://doi.org/10.1093/rheumatology/keae198

Maciel-Fiuza, M. F., Muller, G. C., Campos, D. M. S., do Socorro Silva Costa, P., Peruzzo, J., Bonamigo, R. R., Veit, T., & Vianna, F. S. L. (2023). Role of gut microbiota in infectious and inflammatory diseases. *Front Microbiol*, *14*, 1098386. https://doi.org/10.3389/fmicb.2023.1098386

McArdle, A., Pennington, S., & FitzGerald, O. (2018). Clinical Features of Psoriatic Arthritis: a Comprehensive Review of Unmet Clinical Needs. *Clin Rev Allergy Immunol*, *55*(3), 271-294. https://doi.org/10.1007/s12016-017-8630-7

Mekhail, C., Chouk, M., Prati, C., Wendling, D., & Verhoeven, F. (2020). Prognostic factors of good response to DMARDs in psoriatic arthritis: a narrative review. *Expert Rev Clin Pharmacol*, *13*(5), 505-519.

Michielsens, C. A., den Broeder, N., van den Hoogen, F. H., Mahler, E. A., Teerenstra, S., van der Heijde, D., Verhoef, L. M., & den Broeder, A. A. (2022). Treat-to-target dose reduction and withdrawal strategy of TNF inhibitors in psoriatic arthritis and axial spondyloarthritis: a randomised controlled non-inferiority trial. *Ann Rheum Dis*, *81*(10), 1392-1399.

Orbai, A. M., Tillett, W., Grieb, S., Peterson, S., Holdsworth, E. A., Booth, N., Chakravarty, S. D., & Gossec, L. (2023). Impact of Physician-Defined Flares on Quality of Life and Work Impairment: An International Survey of 2238 Patients With Psoriatic Arthritis. *J Rheumatol*, *50*(1), 76-83. https://doi.org/10.3899/jrheum.211302

Pedraz-Petrozzi, B., Neumann, E., & Sammer, G. (2020). Pro-inflammatory markers and fatigue in patients with depression: A case-control study. *Scientific Reports*, *10*(1), 9494. https://doi.org/10.1038/s41598-020-66532-6

Queiro, R., Sarasqueta, C., Belzunegui, J., Gonzalez, C., Figueroa, M., & Torre-Alonso, J. C. (2002). Psoriatic spondyloarthropathy: a comparative study between HLA-B27 positive and HLA-B27 negative disease. *Semin Arthritis Rheum*, *31*(6), 413-418.

Sandoval, D. C., & Fernández-Ávila, D. G. (2023). Assessment tools in psoriatic arthritis: A review. *Revista Colombiana de Reumatología (English Edition)*, *30*, S75-S86. https://doi.org/https://doi.org/10.1016/j.rcreue.2022.11.003

Schneider, S. (2013). The International Standard Classification of Education 2011. *Comparative Social Research*, *30*, 365-379. https://doi.org/10.1108/s0195-6310(2013)0000030017

Snoeck Henkemans, S. V. J., de Jong, P. H. P., Luime, J. J., Kok, M. R., Tchetverikov, I., Kasiem, F. R., Welby, S., Prickett, A. R., van der Helm-van Mil, A. H. M., Vis, M., & Cicero. (2022). Importance of quick attainment of minimal disease activity for a positive impact on lives of patients with psoriatic arthritis. *RMD Open*, *8*(2).

Tarannum, S., Leung, Y.-Y., Johnson, S. R., Widdifield, J., Strand, V., Rochon, P., & Eder, L. (2022). Sex- and gender-related differences in psoriatic arthritis. *Nature Reviews Rheumatology*, *18*(9), 513-526. https://doi.org/10.1038/s41584-022-00810-7

Tillett, W., Merola, J. F., Thaçi, D., Holdsworth, E., Booth, N., Lobosco, L. S., Milligan, G., Hufford, M. M., Birt, J. A., & Boehncke, W. H. (2020). Disease Characteristics and the Burden of Joint and Skin Involvement Amongst People With Psoriatic Arthritis: A Population Survey. *Rheumatol Ther*, *7*(3), 617-637.

Tripathy, S., Marsland, A. L., Kinnee, E. J., Tunno, B. J., Manuck, S. B., Gianaros, P. J., & Clougherty, J. E. (2021). Long-Term Ambient Air Pollution Exposures and Circulating and Stimulated Inflammatory Mediators in a Cohort of Midlife Adults. *Environ Health Perspect*, *129*(5), 57007.

van Hal, T. W., Mulder, M. L. M., Wenink, M. H., & Vriezekolk, J. E. (2022). Determinants of work and social participation in patients with psoriatic arthritis in the Netherlands: an observational study. *BMC Rheumatol*, *6*(1), 49.

Veale, D. J., & Fearon, U. (2018). The pathogenesis of psoriatic arthritis. *Lancet*, *391*(10136), 2273-2284. https://doi.org/10.1016/S0140-6736(18)30830-4

Walsh, J. A., Ogdie, A., Michaud, K., Peterson, S., Holdsworth, E. A., Karyekar, C. S., Booth, N., Middleton-Dalby, C., Chakravarty, S. D., Dennis, N., & Gossec, L. (2023). Impact of key manifestations of psoriatic arthritis on patient quality of life, functional status, and work productivity: Findings from a real-world study in the United States and Europe. *Joint Bone Spine*, *90*(3), 105534.

Appendix I Supplementary material for retrospective data analysis

I.1 Retrospective data analysis: Supplementary materials for objective 3 (Methods)

Bayesian Regression: Extended Methods

Continuous Outcome Model: PASDAS

The PASDAS is a continuous, composite index of disease activity derived from clinical assessments. For this model:

- The dependent variable Y was the PASDAS score. The "Change in disease score from baseline" (ΔPASDAS) was used as the outcome of the Bayesian Model, Y_{Δ}
- Prior to modeling, PASDAS values were z-score normalized (mean-centered, unit variance) across the dataset.

The generative model for PASDAS was:

$$\mathrm{Y}_{\Delta,\mathrm{i,t}} \sim \mathrm{StudentT}\left(\nu, \mu_{\mathrm{i,t}}, \sigma_{\mathrm{g(i)}}\right)$$

The mean model is given by:

$$\mu_{it} = \beta_0 + \beta_{\text{time}} \cdot t_{it} + \beta_{g_i}^{(\text{group})} + \beta_{g_i}^{(\text{interaction})} \cdot t_{it} + u_i$$

$$u_i \sim \mathcal{N}(0, \tau^2)$$

$$g_i \in \{0\text{:Very Early}, 1\text{:Early}, 2\text{:Intermediate}, 3\text{:Late}\}$$

where:

- μ_{it} is the expected PASDAS score change from baseline for patient i at timepoint t.
- β_0 is the global intercept.
- β_{time} is the fixed effect of time (shared across groups).
- $\beta_{g_i}^{(\text{group})}$ is the group-specific baseline offset, where $g_i \in \{0,1,2,3\}$ corresponds to Very Early, Early, Intermediate, and Late diagnosis.
- $\beta_{g_i}^{(\text{interaction})}$ is the group-specific time slope.
- t_{it} is the scaled time (in months) for patient I at visit t.
- $u_i \sim \mathcal{N}(0, \tau^2)$ is a patient-specific random intercept.

Priors and Missing Data

We employed weakly informative priors throughout the model to provide regularization while retaining flexibility. All fixed-effect coefficients, including the global intercept time effect, group-specific offsets and interaction terms, were assigned Normal priors with mean 0 and standard deviation 1. These priors support moderate effect sizes while preventing overfitting. Random intercepts for each patient were modeled as $\mathcal{N}(0, \tau^2)$ with the group-level standard deviation $\tau \sim \text{HalfCauchy}(2.5)$, a standard weakly informative choice for hierarchical variance parameters. The residual standard deviation σ for each diagnosis group was also given a HalfCauchy prior. To accommodate potential outliers in clinical outcomes, we used a Student-t likelihood with degrees of freedom $\nu \sim \text{Exponential}(1.0)$ allowing for heavy-tailed behavior.

Missing outcome values (Y) were handled natively by the probabilistic programming framework: any observations with missing values were automatically excluded from the likelihood without requiring imputation, preserving uncertainty. We modeled outcome changes from baseline (e.g., PASDAS deltas), and timepoints with entirely missing values were omitted during preprocessing. This approach assumes missingness at random and avoids ad hoc deletion or imputation strategies that could bias posterior estimates.

Categorical-like Outcome Model: MDA

MDA is a composite clinical outcome indicating whether a patient meets a low disease activity threshold. Unlike continuous scores, MDA was modeled on its original scale, without normalization, to preserve interpretability of its categorical nature.

MDA Model: Adaptation for Binary Outcomes
Because of its binary nature, the MDA model required a distinct modeling strategy from that used for continuous outcomes such as PASDAS.

We modeled MDA using a Bayesian logistic regression framework. The linear predictor included the same structure as in the PASDAS model: a global intercept (β_0), a shared time slope (β_{time}), diagnosis group-specific offsets $\left(\beta^{(\text{group})}\right)$ interaction terms $\left(\beta^{(\text{interaction})}\right)$, and random intercepts per patient $(u_i \sim \mathcal{N}(0, \tau^2))$However, instead of using a Student-t likelihood, the model employed a Bernoulli likelihood with a logistic link function:

$$\text{logit}(p_{it}) = \mu_{it} = \beta_0 + \beta_{\text{time}} \cdot t_{it} + \beta_{g_i}^{(\text{group})} + \beta_{g_i}^{(\text{interaction})} \cdot t_{it} + u_i$$

$$Y_{it} \sim \text{Bernoulli}(p_{it})$$

where Y_{it} is the observed binary MDA score for patient i at time t, and p_{it} is the posterior probability of achieving MDA. The probabilities p_{it} were estimated via a logistic transformation $p_{it} = \text{sigmoid}(\mu_{it})$ and stored as a deterministic variable for downstream interpretation.

Differences from the PASDAS Model
Key differences from the continuous PASDAS model include:

- Likelihood: The PASDAS model used a Student-t likelihood to allow for continuous, heavy-tailed residuals, while the MDA model used a Bernoulli likelihood suitable for binary outcomes.
- Link function: PASDAS used an identity link (direct modeling of μ\muμ), whereas MDA used a logit link, transforming the linear predictor to a probability scale.
- Outcome preprocessing: PASDAS scores were z-scored and modeled as changes from baseline, while MDA values were left on the original 0/1 scale. No normalization or delta transformation was applied.
- Inference target: For MDA, posterior estimates represent probabilities of success (MDA achieved), which can be used to assess group-level differences in disease control over time.

As in the PASDAS model, we applied weakly informative Normal priors to all regression coefficients, and a Half-Cauchy prior to the scale of random intercepts. Missing outcome values were excluded from the likelihood automatically, without imputation.

Inference and Model Evaluation
All models were implemented in PyMC (v5) using the No-U-Turn Sampler (NUTS) for posterior inference. Sampling settings included:

- 4 chains,
- 2,000 posterior samples per chain,
- 500 warm-up iterations,
- Target acceptance probability of 0.9,
- Maximum tree depth of 12.

Posterior diagnostics (R-hat, trace plots) confirmed convergence (R-hat < 1.01 for all parameters). Posterior predictive checks were performed to validate model fit against observed distributions and temporal trends.

Constructing the kNN Feature

To construct a predictive risk score from the training data without introducing data leakage, we implemented an out-of-fold (OOF) strategy:

1. The dataset was randomly split into 5 folds using
2. For each fold:
    - A kNN regressor (with k=30 neighbors) was trained on only the uncensored (event occurred) patients in the training folds, using standardized covariates X.
    - The model predicted expected time-to-event for patients in the validation fold.
    - These predictions were converted to risk scores using the transformation $\text{Risk} = \frac{1}{\text{predicted time} + \varepsilon}$
3. The resulting risk scores were concatenated to form a complete OOF vector, which was added as a new feature to the Cox model input.

This procedure ensured that each patient's kNN prediction was generated from a model trained on a disjoint subset of the data, thus preventing any information leakage from future outcomes.

After constructing the OOF kNN risk scores, the same covariates X and the newly added feature were used to fit a Cox proportional hazards model targeting the same phase-specific event as the kNN model. This design ensured conceptual alignment: the kNN feature predicted time-to-event for reaching a given EULAR phase, based on the same covariates used in the survival model.

Several precautions were implemented to prevent data leakage and ensure robust estimation:

- No neighbor from the test set was used to construct the reference set, preserving strict temporal and sampling independence.
- kNN calculations were repeated separately for each treatment phase model, using time-to-event features and outcome definitions tailored to the respective endpoint (e.g., reaching Phase 2 vs. Phase 3).
- Outcome values used for aggregation were hidden for test patients and only obtained from the training patients.
- Standardization and preprocessing were done exclusively on training data, and test features were transformed using these same training parameters.

This design ensures that the kNN feature provides an honest, generalizable estimate of escalation risk based on historical patterns in clinically similar patients, without inadvertently encoding knowledge about the test set.

I.2 Retrospective data analysis: Supplementary materials for objective 1 (Results)

Appendix Table 1 Descriptions of educational category corresponding to each International Standard Classification of Education (ISCED) level, as used to classify participants' highest completed level of education (Schneider, 2013).

ISCED Level	Full Description
Level 1	Continued education to primary school started but not finished
Level 1	Continued education to only received primary school education
Level 2	Continued education to educated to secondary school level
Level 3	Continued education to sixth form
Level 3	Continued education to highest grade: Twelve or GED
Level 4	Continued education to received vocational training
Level 5	Continued education to received further education
Level 5	Continued education to received education at technical college
Level 6	Continued education to highest grade: College graduate

Appendix Table 2 Baseline ethnicity distribution by cohort.

	T0		Missing %		p-value
Outcomes	**DEPAR (N=936)**	**Reuma.pt (N=753)**	**DEPAR (N=936)**	**Reuma.pt (N=753)**	
Ethnicity	-	-	0%	0%	NA
Dutch	100%	NA	-	-	-
Black	NA	0.27%	-	-	-
Mulatto	NA	0.27%	-	-	-
White non-EU	NA	1.2%	-	-	-
Asia SE	NA	0.13%	-	-	-
Unknown	NA	31.74%	-	-	-
White EU	NA	65.87%	-	-	-
Asian IND	NA	0.4%	-	-	-
Jewish	NA	0.13%	-	-	-

I.3 Retrospective data analysis: Supplementary materials for objective 2 (Results)

Appendix Table 3 Baseline ethnicity distribution by sex.

	T0		Missing %		p-value
Outcomes	**Female (N=810)**	**Male (N=879)**	**Female (N=810)**	**Male (N=879)**	
Ethnicity	-	-	0%	0%	NA
Dutch	59.26%	51.88%	-	-	-
Black	0.25%	0.00%	-	-	-
Mulatto	0.25%	0.00%	-	-	-
White non-EU	0.25%	0.80%	-	-	-
Asia SE	0.00%	0.11%	-	-	-
Unknown	11.98%	16.15%	-	-	-
White EU	28.02%	30.60%	-	-	-
Asian IND	0.00%	0.34%	-	-	-
Jewish	0.00%	0.11%	-	-	-

Appendix Table 4 Mixed linear model estimates for longitudinal DAPSA scores.

Term	Coefficient	Std. Error	z	p-value	95% CI
Intercept	15.034	1.009	14.91	<0.001	[13.058, 17.011]
Male	-2.673	0.576	-4.64	<0.001	[-3.803, -1.543]
Reuma.pt	2.609	0.544	4.8	<0.001	[1.544, 3.675]
Time	-0.203	0.018	-11.23	<0.001	[-0.239, -0.168]
Male × Time	-0.019	0.025	-0.77	0.443	[-0.069, 0.030]
Age	0.021	0.019	1.12	0.261	[-0.016, 0.058]
Group Var	58.205	0.444			
Note: Model fitted using restricted maximum likelihood (REML). DAPSA was the outcome variable. No. Observations: 5552; No. Groups: 1313; Mean [Min, Max] group size: 4.2 [1,8]. CI: Confidence interval.					

Appendix Table 5 Generalized estimating equations model for MDA.

Term	Coefficient	Std. Error	z	p-value	95% CI	OR [95% CI]
Intercept	-0.494	0.181	-2.731	0.006	[-0.849, -0.139]	0.610 [0.428-0.870]
Male	0.479	0.105	4.566	<0.001	[0.273, 0.684]	1.614 [1.314-1.982]
Reuma.pt	-1.68	0.11	-15.217	<0.001	[-1.897, -1.464]	0.186 [0.150-0.231]
Time	0.028	0.003	8.006	<0.001	[0.021, 0.034]	1.028 [1.022-1.034]

Male × Time	0.018	0.005	3.79	<0.001	[0.009, 0.027]	1.018 [1.008-1.028]
Age	-0.009	0.003	-2.66	0.008	[-0.016, -0.002]	0.991 [0.985-0.997]
Note: Model fitted using a binomial distribution and logit link. MDA was the outcome variable. No. Observations: 7488, No. Groups: 1562, Mean [Min, Max] group size: 4.8 [1,8]; OR: Odds ratio.						

Appendix Table 6 Mixed linear model estimates for longitudinal VAS pain scores.

Term	Coefficient	Std. Error	z	p-value	95% CI
Intercept	47.35	2.27	20.85	<0.001	[42.90, 51.80]
Male	-9.44	1.29	-7.35	<0.001	[-11.96, -6.92]
Reuma.pt	0.61	1.22	0.5	0.616	[-1.79, 3.01]
Time	-0.29	0.04	-8.16	<0.001	[-0.36, -0.22]
Male × Time	-0.04	0.05	-0.73	0.463	[-0.13, 0.06]
Age	-0.06	0.04	-1.31	0.190	[-0.14, 0.03]
Group Var	336.54	1.03			
Note: Model fitted using restricted maximum likelihood (REML). VAS pain was the outcome. No. Observations: 5963; No. Groups: 1342; Mean [Min, Max] group size: 4.4 [1,8].					

Appendix Table 7 Mixed linear model estimates for longitudinal HAQ scores.

Term	Coefficient	Std. Error	z	p-value	95% CI
Intercept	0.45	0.05	8.29	<0.001	[0.34, 0.55]
Male	-0.28	0.03	-9.54	<0.001	[-0.34, -0.22]
Reuma.pt	0.02	0.03	0.51	0.610	[-0.04, 0.08]
Time	-0.002	0.001	-3.54	<0.001	[-0.003, -0.001]
Male × Time	-0.003	0.001	-3.57	<0.001	[-0.004, -0.001]
Age	0.007	0.001	7.14	<0.001	[0.01, 0.01]
Group Var	0.22	0.04			
Note: Model fitted using restricted maximum likelihood (REML). HAQ was the outcome variable. No. Observations: 5931; No. Groups: 1296; Mean [Min, Max] group size: 4.6 [1,8].					

Appendix Table 8 Medication use over time by class and sex. Proportions of patients (%) treated with each medication class at each visit are shown, alongside total group sizes (N). F: Female; M: Male;.

	N	None %	GC %	MTX %	csDMARD %	bDMARD %	tsDMARD %
T0	F (N=810)	45.06	15.06	41.11	8.52	7.78	0.12
	M (N=879)	45.85	12.51	43	7.28	8.42	0

T3	F (N=810)	38.15	11.23	47.28	12.47	14.32	0
	M (N=879)	35.27	11.26	52.56	10.69	14.9	0
T6	F (N=686)	28.43	12.24	52.33	16.33	19.68	0.15
	M (N=750)	25.47	11.33	58.67	12.8	21.2	0
T9	F (N=656)	28.51	11.74	50.3	17.07	22.87	0
	M (N=708)	24.72	10.88	58.19	12.29	23.87	0
T12	F (N=622)	27.17	10.93	49.52	18.01	26.21	0.16
	M (N=675)	24.59	10.67	57.78	13.78	26.52	0
T18	F (N=542)	25.09	10.52	48.89	19.56	30.44	0.74
	M (N=584)	23.29	9.76	56.85	14.9	28.25	0.17
T24	F (N=475)	28	10.95	46.95	18.53	30.74	0.63
	M (N=513)	23.39	9.16	56.92	12.09	29.82	0.19
T36	F (N=349)	41.26	8.31	36.96	15.47	29.23	0.29
	M (N=375)	30.4	8.27	48.8	11.73	30.67	0.8

Appendix Table 9 Retention rates at 6, 12, 24, and 36 months for each medication class by sex along with number of patients per group and p-values from log-rank tests. Values are shown as percentages with 95% confidence intervals.

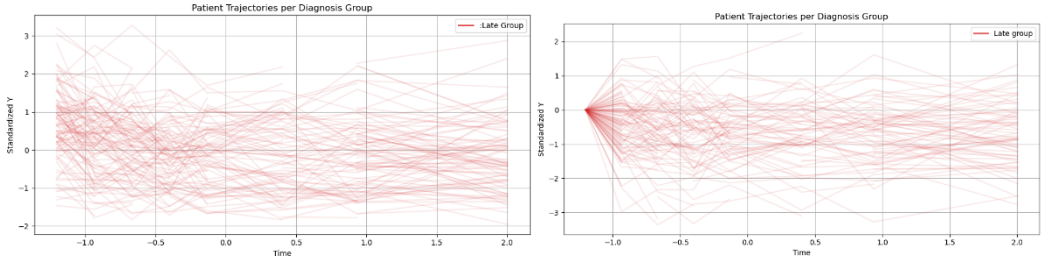
		GC	**MTX**	**csDMARD**	**bDMARD**
	F [N]	246	603	286	369
	M [N]	210	661	215	339
	p-value	0.007	<0.001	0.156	0.001
M6	**F**	44% [37%, 50%]	78% [74%, 81%]	64% [58%, 70%]	82% [78%, 86%]
	M	58% [51%, 65%]	84% [81%, 87%]	69% [62%, 75%]	90% [86%, 93%]
M12	**F**	34% [28%, 41%]	63% [59%, 67%]	46% [40%, 52%]	66% [60%, 71%]

	M	47% [40%, 54%]	76% [72%, 79%]	54% [47%, 61%]	77% [72%, 81%]
M24	**F**	28% [22%, 34%]	51% [46%, 55%]	31% [25%, 37%]	48% [42%, 54%]
	M	36% [28%, 43%]	61% [57%, 65%]	35% [27%, 42%]	58% [51%, 64%]
M36	**F**	22% [16%, 29%]	41% [36%, 46%]	24% [17%, 31%]	38% [31%, 46%]
	M	31% [23%, 39%]	51% [46%, 56%]	27% [20%, 35%]	49% [41%, 57%]

Absolute vs. Relative Change in Disease Activity: Choosing the disease activity indicator

Appendix Figure 1 Absolute PASDAS scores (left panel) and changes from baseline (ΔPASDAS; right panel)

Appendix Figure 1 compares patient-level PASDAS trajectories using two alternative outcome definitions: absolute scores (left panel) and changes from baseline (ΔPASDAS; right panel), both standardized (data from DEPAR cohort). While absolute PASDAS trajectories show considerable within-group heterogeneity with no clear temporal pattern, the ΔPASDAS formulation reveals a more coherent structure. All patient ΔPASDAS trajectories begin at a common baseline (Δ = 0), reducing between-subject variability at time zero and highlighting meaningful temporal changes. This anchoring enables the model to focus on within-patient response rather than baseline differences, enhancing the detection of group-level effects. Additionally, the divergence of trajectories over time in the ΔPASDAS plot reveals stronger temporal signals than in the raw PASDAS plot, where baseline heterogeneity masks patterns.

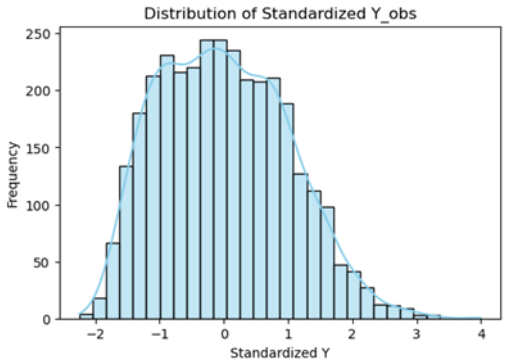


The clearer temporal alignment and group-level variability in the right panel support the choice of ΔPASDAS as the response variable in the Bayesian regression model.

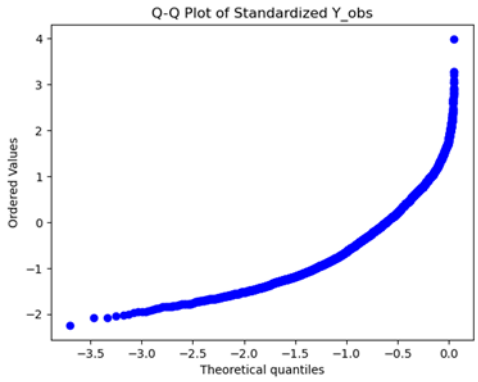
Appendix Figure 1 Absolute PASDAS scores (left panel) and changes from baseline (ΔPASDAS; right panel)

Differences in disease activity in Late vs Early Diagnosis groups: PASDAS (Extended Results)

The distribution of the standardized ΔPASDAS scores (Appendix Figure 2) approximated normality, though the Q-Q plot (Appendix Figure 3) demonstrates that while the central distribution of standardized PASDAS values is approximately normal, the data exhibit heavier tails and positive skewness. These deviations justify the model's use of a Student-t likelihood, which is more robust to outliers and non-normal residuals than a Gaussian alternative.



Appendix Figure 2 Distribution of normalized ΔPASDAS scores



Appendix Figure 3 QQ plot to compare ΔPASDAS distribution to Normal

Reuma.pt population comparisons at Clinical Visits, from Baseline to 36-month visit

We present here results on the outcomes available on Reuma.pt cohort (DAPSA and SF36). Due to high missingness (see Table 7 and Table 8, the sample sizes per clinical visit and diagnosis group ranged from n = 20–30 for SF-36 domains and n = 70–100 for DAPSA), few comparisons reach statistical significance. However, overall trends can be observed and comparisons with DEPAR cohort can be made.

Group-mean disease activity and quality of life scores over time in Early (<1 year) vs Late (>1 year) diagnosis groups in the Reuma.pt cohort are shown in Appendix Table 10. Values represent group means at each visit timepoint (First Visit, T3, T6, T9, T12, T18, T24, T36) for disease activity indices (DAPSA) and SF-36 health-related quality of life domains. Early diagnosis group scores are shown in green when the difference with the Late group is statistically significant ($p < 0.05$), based Mann-Whitney U and t tests using group-level summary statistics. Normality assumption assessed by Lilliefors test. Bold font comparisons using t-test, normal font comparisons using Mann-Whitney-U test.

Appendix Table 10 Group-mean disease activity and quality of life scores over time in Early (<1 year) vs Late (>1 year) diagnosis groups in the Reuma.PT cohort. The Reuma.pt cohort analysis focused on longitudinal comparisons of disease activity and health-related quality of life (HRQoL) across early and late diagnosis groups. The results presented reflect only those variables found to be statistically significant at the $p < 0.05$ level.

Measure \ Visit	First Visit	T3	T6	T9	T12	T18	T24	T36
DAPSA	Early 27.56, Late 25.68	Early 13.51, Late 16.06	Early 11.20, Late 13.01	Early 10.87, Late 14.31	Early 9.55, Late 11.69	Early 9.65, Late 11.69	Early 9.81, Late 10.55	Early 9.62, Late 9.97
sf36 vitality	Early 38.48, Late 51.32	Early 56.44, Late 48.24	Early 57.69, Late 51.12	Early 59.33, Late 52.83	Early 61.46, Late 45.81	Early 58.81, Late 45.89	Early 57.76, Late 48.63	Early 59.40, Late 49.83
sf36 mental	Early 58.53, Late 60.55	Early 66.74, Late 64.68	Early 73.29, Late 60.02	Early 76.14, Late 65.59	Early 75.07, Late 68.76	Early 72.26, Late 68.05	Early 64.16, Late 60.13	Early 70.49, Late 61.50
sf36 general health	Early 42.50, Late 50.82	Early 50.74, Late 46.90	Early 54.31, Late 44.65	**Early 58.57, Late 40.63**	Early 54.58, Late 47.39	Early 50.51, Late 41.60	Early 48.21, Late 39.77	Early 53.11, Late 45.05
sf36 role physical	Early 28.91, Late 45.31	Early 58.42, Late 52.48	Early 53.37, Late 58.75	Early 71.88, Late 60.42	Early 63.54, Late 59.38	Early 66.96, Late 57.25	Early 60.53, Late 55.71	Early 64.06, Late 56.03
sf36 bodily pain	**Early 30.21, Late 50.18**	Early 50.11, Late 46.70	Early 48.15, Late 51.65	Early 66.79, Late 54.68	Early 59.28, Late 47.72	Early 55.31, Late 50.35	Early 55.79, Late 49.60	**Early 66.91, Late 49.84**
sf36 role emotional	Early 45.08, Late 41.18	Early 67.03, Late 64.94	Early 66.67, Late 74.58	Early 69.64, Late 78.95	Early 73.15, Late 74.86	Early 73.02, Late 72.67	Early 67.54, Late 66.20	Early 76.56, Late 66.07
sf36 social functioning	Early 55.65, Late 63.59	Early 70.20, Late 68.53	Early 83.65, Late 68.88	Early 86.71, Late 69.79	Early 82.72, Late 65.66	Early 74.57, Late 70.59	Early 77.16, Late 70.94	Early 81.38, Late 68.03
sf36 physical functioning	**Early 40.62, Late 60.59**	Early 59.35, Late 58.65	Early 58.50, Late 60.78	Early 82.30, Late 62.37	Early 66.60, Late 63.33	Early 67.74, Late 60.42	Early 69.21, Late 49.20	Early 68.91, Late 62.06

Due to smaller sample sizes in Reuma.pt compared to prior cohorts such as DEPAR, fewer statistically significant results were observed. Specifically, the sample sizes per clinical visit and diagnosis group ranged from n = 20–30 for SF-36 domains and n = 70–100 for DAPSA scores, whereas DEPAR included over 200 samples per group and time point. This difference limits statistical power but does not diminish the importance of the trends observed, which are directionally consistent with DEPAR findings.

DAPSA
DAPSA was the only disease activity score calculated in Reuma.pt (PASDAS and GRACE were not included due to missingness of some variables required to calculate these scores). At the T12 visit, DAPSA scores were significantly lower in the early diagnosis group (10.87) compared to the late group (14.31), suggesting more effective disease control when intervention occurs earlier in the disease course.

Health-Related Quality of Life (SF-36 Domains)
Several SF-36 subdomains reached statistical significance at isolated time points, highlighting the selective but consistent benefits of early diagnosis:

- Vitality: Statistically higher scores were observed in the early group at T18 (61.46 vs. 45.81) and T24 (58.81 vs. 45.89), reflecting sustained energy and mental well-being benefits from early intervention.
- General Health: At T12, the early group reported significantly better general health perceptions (58.57 vs. 40.63), consistent with improved disease management.
- Bodily Pain: At baseline, the early group reported significantly less pain (30.21 vs. 50.18), and although the trend reversed slightly by T36 (66.91 vs. 49.84), the initial benefit suggests early pain mitigation.
- Social Functioning: At T18, social functioning scores were significantly higher in the early group (82.72 vs. 65.66), suggesting improved social engagement and participation.
- Physical Functioning: Significant group differences emerged at multiple time points, including the first visit (40.62 vs. 60.59), T12 (82.30 vs. 62.37), and T36 (69.21 vs. 49.20), reinforcing a pattern of superior physical capacity in the early group.

While not all SF-36 domains reached significance—likely due to limited statistical power—the consistency in direction across domains and time points mirrors trends seen in the DEPAR cohort and supports the general hypothesis that early diagnosis is associated with improved outcomes.

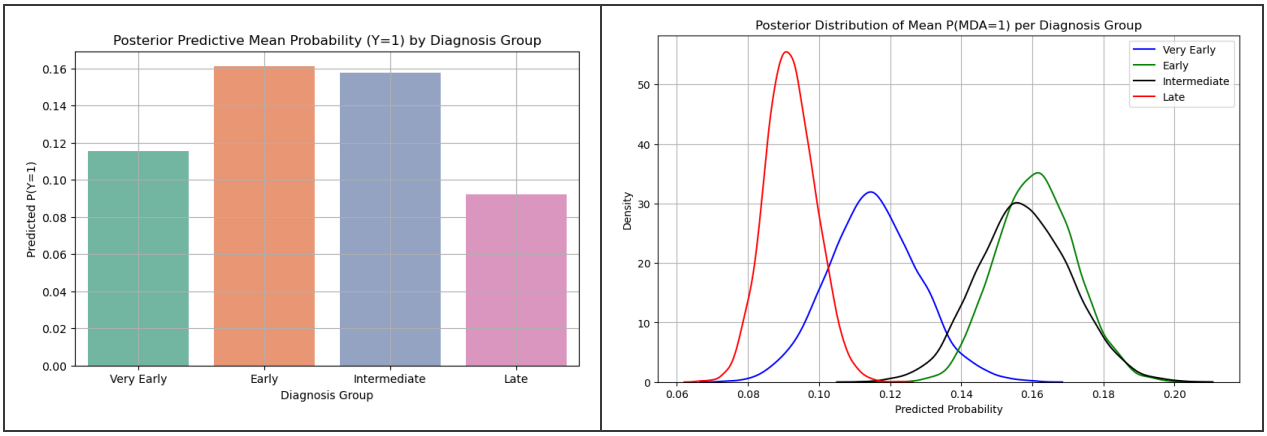
Highlight
Despite small sample sizes, significant differences between early and late diagnosis groups were observed in DAPSA and multiple SF-36 domains (e.g., vitality, general health, and physical functioning), reinforcing the impact of early diagnosis on long-term disease control and quality of life in PsA patients.

Differences in disease activity in Late vs Early Diagnosis groups in Reuma.pt cohort

We repeat the Bayesian regression analysis the on MDA in Reuma.pt cohort. The selection of more severe PsA cases in Reuma.pt (patients initiating biologic therapy), is reflected to overall lower probability of achieving MDA in reuma.pt, compared to the MDA probabilities on the combined reuma.pt and DEPAR population (Figure 14). This result is aligned with the worst disease outcomes in Reuma.pt patients observed in Figure 7 (Linear Mixed Effects model on MDA, Reuma.pt Odds Ratio < 1).

Appendix Figure 4 presents the posterior predictive mean probabilities of achieving MDA by diagnosis group (left), along with the posterior distribution of these group-level means (right). The early diagnosis groups ("Very Early" and "Early") again show higher predicted probabilities of MDA compared to the "Late" group. However, the absolute probabilities are markedly lower than those observed in the NL cohort.

Despite the overall lower predicted probabilities in the Reuma.pt cohort, the group-level structure still reveals important distinctions. As shown in Appendix Figure 4 and summarized in Appendix Table 11, the Late diagnosis group remains consistently and clearly separated from the Early and Intermediate groups, with a significantly lower probability of achieving MDA. Interestingly, the Intermediate group does not follow the expected downward trajectory from the Early group—instead, their outcomes are nearly identical, suggesting that the treatment response in this group is not measurably worse than in Early-diagnosed patients.



Appendix Figure 4 (Left) Posterior predictive mean probability of achieving Minimal Disease Activity (MDA) by diagnosis group. (Right) Posterior distributions of the group-level mean probability of MDA.
Bars represent the group-level average predicted probability. Each curve reflects the uncertainty in P(MDA=1) for a diagnosis group, sampled from the Bayesian posterior

Appendix Table 11 Posterior estimates of the mean probability of achieving Minimal Disease Activity (MDA) by diagnosis group and 95% Highest Density Intervals (HDI)

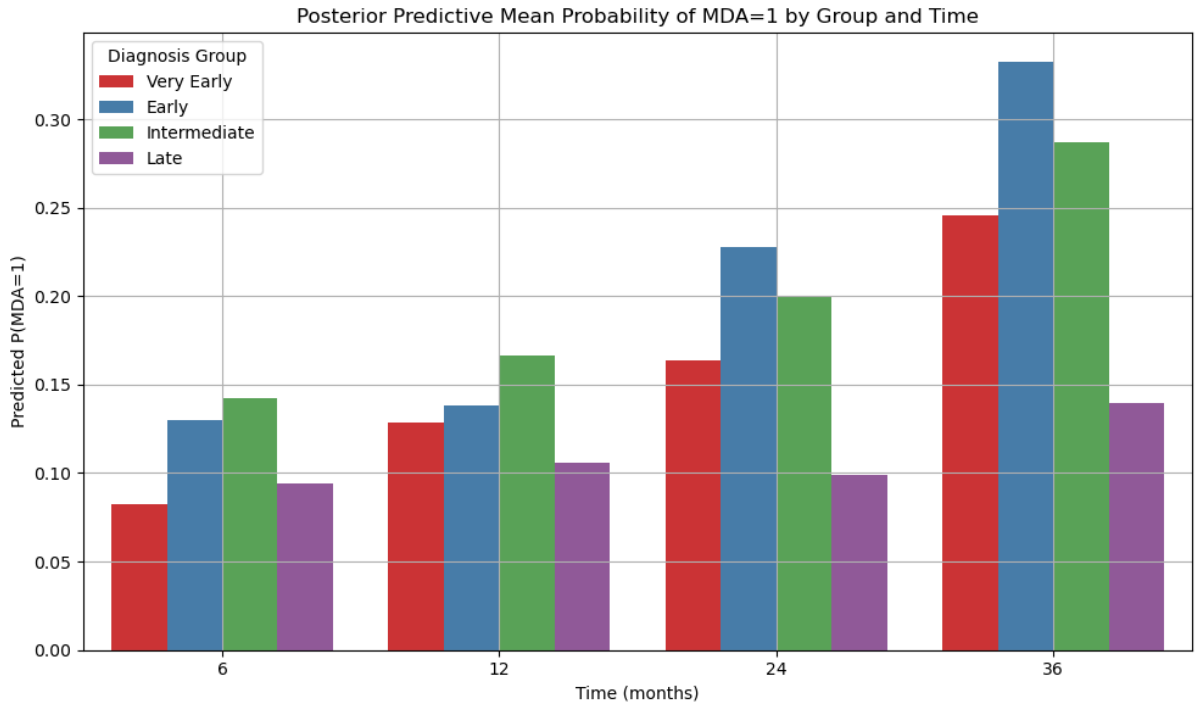
Diagnosis Group	Mean P(MDA=1)	95% HDI
Very Early	0.115	[0.091, 0.142]
Early	0.161	[0.139, 0.182]
Intermediate	0.158	[0.133, 0.184]
Late	0.092	[0.078, 0.106]

A notable deviation from the expected gradient is seen in the Very Early group, which shows the lowest predicted MDA probabilities after the Late group. This anomaly could be explained by cohort-specific selection effects: patients classified as Very Early in Reuma.pt may include individuals whose disease severity prompted early escalation to biologic therapy, thus skewing this group toward more severe cases despite earlier diagnosis. This highlights the nuanced nature of real-world clinical cohorts like Reuma.pt, where diagnostic timing may intersect with severity in complex ways.

Appendix Figure 5 explores the temporal evolution of MDA probability by group across clinical follow-ups. In both cohorts, MDA probability increased over time for all groups, suggesting that continued treatment exerts beneficial effects. However, in Reuma.PT the absolute gains were more modest, and group separations—though present—were less pronounced over time. The

Late group consistently lagged behind, reinforcing the finding that diagnostic delay is associated with poorer long-term disease control.

Taken together, these results replicate the diagnostic timing effect found in the NL cohort but highlight the challenges of achieving MDA in more severe PsA cases. The Bayesian model again demonstrates a robust association between early diagnosis and improved clinical outcomes, even within a more treatment-refractory population. The reduced absolute MDA rates in Reuma.pt reflect the more severe disease trajectory in patients treated with biologics.



Appendix Figure5 Posterior predictive mean probability of achieving Minimal Disease Activity (MDA) by diagnosis group and time point.

Highlight

Bayesian modeling showed that patients diagnosed late had substantially lower probabilities of achieving MDA, with the Late group clearly separated from the Early and Intermediate groups. This underscores a persistent diagnostic timing effect, even in a biologics-treated, high-severity population.