Delft University of Technology

Document Version
Final published version

Licence
CC BY-NC

Citation (APA)
Maarseveen, T., Nagafuchi, Y., Lend, K., Glas, H. K., Gröndal, G., Gudbjornsson, B., Haavardsholm, E. A., Reinders, M. J. T., Knevel, R., & More Authors (2026). Patients Improve, Patterns Persist: Longitudinal Stability of Rheumatoid Arthritis Joint Involvement Patterns. *Arthritis Care and Research*. https://doi.org/10.1002/acr.80174

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Check for updates

Arthritis Care & Research
Vol. 0, No. 0, Month 2026, pp 1–12
DOI 10.1002/acr.80174
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

American College of Rheumatology
Empowering Rheumatology Professionals

Patients Improve, Patterns Persist: Longitudinal Stability of Rheumatoid Arthritis Joint Involvement Patterns

Tjardo Maarseveen, MSc,[1] Yasuo Nagafuchi, MD,[1,2] Kristina Lend, MSc,[3,4] Seyam Alemyar, MSc,[1] Herman Kasper Glas, MD,[5] Gerdur Gröndal, MD,[6,7] Bjorn Gudbjornsson, MD,[6,7] Espen A. Haavardsholm, MD,[8] Kim Hørslev Petersen, MD,[9,10] Jon Lampa, MD,[4,11] Merete Lund Hetland, MD,[12,13] Dan Nordström, MD,[14,15] Mike Nurmohamed, MD,[3,16] Mikkel Østergaard, MD,[12,13] Anna Rudin, MD,[17,18] Marte Schrumpf Heiberg, MD,[8] Tuulikki Sokka-Isler, MD,[19,20] Till Uhlig, MD,[8,21] Josien van Dieren, MD,[22] Ronald van Vollenhoven, MD,[3,4] Tom Huizinga, MD,[1] Marcel J. T. Reinders, PhD,[23,24] Sytske Anne Bergstra, PhD,[1] Erik van den Akker, PhD,[23,24] and Rachel Knevel, MD[1,23,25]

Objective. Rheumatoid arthritis is a heterogeneous disease. Data-driven approaches, from synovial histology to joint involvement patterns (JIPs), have sought to define clinically meaningful subgroups. Whether these subgroups represent stable phenotypes or transient disease states remains unclear. We aimed to develop a novel analytical method to measure longitudinal stability of these subsets.

Methods. We analyzed two randomized controlled trials, Nordic Rheumatic Diseases Strategy Trials and Registries (NORD-STAR; n = 630) and Behandel-Strategieën (BeST; n = 445), and one real-world cohort (Reumazorg Zuidwest Nederland [RZWN], n = 252), all treated with disease-modifying antirheumatic drugs under a treat-to-target strategy. Patients were assigned to previously defined JIP subgroups based on shared clinical features. Disease progression was then tracked over one year along the two principal axes of JIP variation: overall severity (oligoarticular vs polyarticular) and anatomic distribution (hand vs foot). Longitudinal stability was assessed by measuring how phenotypically similar each patient remained to others sharing their initial joint pattern and was quantified using silhouette-like scores.

Results. Effective treatment led to convergence toward mild disease. Nevertheless, patients sharing similar baseline JIPs maintained clustering together over time significantly more than expected by chance in BeSt (P = 0.018–0.038), NORD-STAR (P = 0.004–0.017), and RZWN (P < 0.001), except for J polyarthritis in BeSt and JIP in oligoarthritis in NORD-STAR. When stratifying by anatomic region, we found that improvement in foot involvement lagged behind hand involvement across trials.

Conclusion. Patients with similar baseline JIPs maintain phenotypic similarity over time, suggesting JIP subgroups reflect distinct, stable phenotypes rather than transient disease states. Our tool is available to study the longitudinal stability of any subset.

INTRODUCTION

Rheumatoid arthritis (RA) is a heterogeneous disease, prompting longstanding debate about whether it represents a collection of distinct subdiseases rather than a single entity.[1] Most recent stratification efforts have used a data-driven approach, focusing on histopathological features of the synovium to define patient subsets.[2,3] Similarly, we also conducted a data-driven

Supported by Horizon Europe (grant 101095052 [to SQUEEZE Consortium], grant 101080711 [to SPIDeRR], ZonMW Klinische Fellow grant 40-00703-97-19069, and ZonMw Open Competitie 2021 grant 09120012110075).
[1]Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands; [2]Division of Rheumatology and Clinical Immunology, Department of Medicine, Jichi Medical University, Tochigi, Japan; [3]Department of Rheumatology and Clinical Immunology, Amsterdam University Medical Center, Amsterdam, the Netherlands; [4]Division of Rheumatology, Department of Medicine, Center for Molecular Medicine, Karolinska Institute, Stockholm, Sweden; [5]Department of Rheumatology, Amphia Ziekenhuis, Breda, the Netherlands; [6]Centre for Rheumatology Research, Landspitali University Hospital, Reykjavik, Iceland; [7]Faculty of Medicine, University of Iceland, Reykjavik, Iceland; [8]Center for Treatment of Rheumatic and Musculoskeletal Diseases, Diakonhjemmet Hospital, Oslo, Norway; [9]Danish Hospital for Rheumatic Diseases, University Hospital of Southern Denmark, Sønderborg, Denmark; [10]Department of Regional Health Research, University of Southern Denmark, Odense, Denmark; [11]Department of Gastroenterology, Dermatology and Rheumatology, Karolinska University Hospital, Stockholm, Sweden; [12]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; [13]Copenhagen Center for Arthritis Research, Center for Rheumatology and Spine Diseases, Centre for Head and Orthopedics, Rigshospitalet, Glostrup, Denmark;

SIGNIFICANCE & INNOVATION

- We introduce a novel analytical framework for quantifying the longitudinal stability of patient subsets, even in the context of intensive treat-to-target therapy.
- We found that patients with similar joint involvement patterns (JIPs) at baseline were significantly more likely to remain clustered together over time than expected by chance (across three cohorts), supporting the notion that JIPs may reflect stable phenotypes.
- Our findings contribute to a central debate in rheumatology: whether phenotypic differences in rheumatoid arthritis represent transient disease stages or "stable" phenotypes by providing a replicable methodology for studying subgroup stability over time.
- Although joint involvement generally improved, foot joints showed comparatively less improvement than hand joints in the trial cohorts.

approach,[4] but we used routinely collected information from the electronic health records (EHRs), identifying four clinical subsets—defined primarily by their joint involvement patterns (JIPs): foot dominant arthritis (JIP-foot), oligoarthritis (JIP-oligo), hand-dominant arthritis (JIP-hand), and polyarthriticular arthritis (JIP-poly). These anatomic distinctions corresponded to differential treatment responses and distinct synovial tissue characteristics, hinting at different biologic underpinnings. A central question raised by Triaille et al,[5] however, is whether such heterogeneity reflects "stable" phenotypes or merely transient disease states through which all patients may eventually pass.

The treat-to-target paradigm, designed to steer all patients toward mild disease, presents a challenge for studying this subset stability: by effectively reducing the phenotypic differences across patients, intensive therapy may obscure true subgroup differences over time, making it difficult to determine whether observed patterns reflect fixed subtypes or transient disease stages. This complexity motivated the development of a novel temporal dynamics method. In this post hoc study, we identified joint involvement subgroups and tested whether they were longitudinally stable by examining whether patients co-clustered with their initial neighbors regardless of disease progression and choice of treatment. Our central hypothesis was that if baseline JIPs reflect stable underlying biology, patients will remain more similar to their initial neighbors (those with the same phenotype) after one year than to patients with other joint patterns.

MATERIALS AND METHODS

Study cohorts. Three cohorts were analyzed. First, the Behandel-Strategieën (BeST) trial (meaning "treatment strategies") enrolled 508 patients with early RA who were disease-modifying antirheumatic drug (DMARD) naive between 2000 and 2002 (Dutch Trial Register: NTR262 and NTR265).[6] In this trial, RA was defined according to the 1987 American College of Rheumatology (ACR) criteria,[7] with symptom duration of two years or less. Patients were randomized to one of four treatment arms: (1) sequential monotherapy starting with methotrexate (MTX), (2) step-up combination therapy starting with MTX, (3) initial combination therapy with MTX, sulfasalazine, and prednisone, or (4) initial combination therapy with infliximab and MTX. Complete baseline data for clustering[1] were available for 445 patients. Treatment was adjusted according to a treat-to-target strategy based on Disease Activity Scores (DASs; based on 44 joints). If DAS remained >2.4, treatment was intensified. Conversely, if patients achieved sustained low disease activity (DAS < 2.4 for at least six months), treatment was tapered to MTX monotherapy and discontinued entirely if remission was sustained (DAS < 1.6 for at least six months).

The second cohort was the Nordic Rheumatic Diseases Strategy Trials and Registries (NORD-STAR) trial (EudraCT: 2011-004720-35; ClinicalTrials.gov: NCT01491815). The trial enrolled 812 patients between 2012 and 2018, of whom 630 had complete one-year follow-up data[8,9] (Supplementary Figure S1). RA was defined by 2010 ACR/EULAR criteria.[10] Patients were randomly assigned 1:1:1:1 to four arms: (1) conventional therapy with MTX and prednisolone in Sweden, Norway, Iceland, and the Netherlands or triple therapy with sulfasalazine, hydroxychloroquine, and MTX and mandatory intra-articular glucocorticoid injections in Denmark and Finland; (2) MTX plus certolizumab; (3) MTX plus abatacept; or (4) MTX plus tocilizumab.

Finally, we also included the real-world hospital data of the Reumazorg Zuidwest Nederland (RZWN), which featured

[14]Department of Medicine and Rheumatology, Faculty of Medicine, Helsinki University Hospital, Helsinki, Finland; [15]Department of Medicine, University of Helsinki, Helsinki, Finland; [16]Amsterdam Rheumatology and Immunology Center, Reade Amsterdam, Noord-Holland, the Netherlands; [17]Rheumatology Clinic, Sahlgrenska University Hospital, Gothenburg, Sweden; [18]Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy, Gothenburg, Sweden; [19]Rheumatology Clinic, Hospital Nova of Wellbeing Services County of Central Finland, Jyväskylä, Finland; [20]Faculty of Health Sciences, University of Eastern Finland, Kuopio, Finland; [21]University of Oslo, Oslo, Norway; [22]Rheumatology Clinic, Reumazorg Zuid West Nederland, Van Hertumweg, the Netherlands; [23]Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, the Netherlands; [24]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, the Netherlands; [25]Newcastle University School of Clinical Medical Sciences, Newcastle upon Tyne, United Kingdom.

Dr Nagafuchi and Ms Lend contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80174.

Address correspondence via email to Rachel Knevel, MD, at r.knevel@lumc.nl.

Submitted for publication May 18, 2026; accepted in revised form September 24, 2026.

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

515 patients with early RA who were DMARD naive identified from EHRs in the RZWN registry between 2015 and 2022. Follow-up data were available for 210 patients. RA diagnosis was identified from clinical notes using a validated zero-shot strategy with a DeepSeek-R1 large language model (32B parameters), applied within a 1.5-year window from each patient's first visit.[11]

Ethics approval. Ethics approval was obtained from the relevant national and regional ethics committees across all participating countries. Approvals were granted by the following: Regionala etikprövningsnämnden i Stockholm, Sweden (ID: 2011/2069-31/4); Videnskabsetiske Komite for Region Hovedstaden, Denmark (ID: H-2-2013-153); Regional Committees for Medical and Health Research Ethics, Norway (ID: 2014/2191/REC South East); the Ethics Committee of Internal Medicine at Helsinki University Hospital, Finland (ID: 240/13/03/01/2012); Medisch Ethische Toetsingscommissie voor het Slotervaartziekenhuis en Reade, the Netherlands (ID: NL60775.048.17); and the National Bioethics Committee, Iceland (ID: 13-085). Participants provided written informed consent in accordance with the Declaration of Helsinki.

JIP subgroup assignment. To identify the different joint patterns, we applied the Projecting Observations on a Deep Learned Embedding (POODLE) framework, defined in our original study.[12] In that study, a multimodal autoencoder was used to capture phenotypic variation across multiple EHR modalities for 1,387 patients from the Leiden University Medical Center (LUMC).[13] The model incorporated clinical parameters including serology, blood markers, joint involvement (DAS in 66 joints [DAS66]/DAS68 assessment), age, and sex. It learned a low-dimensional representation (ie, a patient embedding) in which JIPs emerged as a key-axis of variation.

Using POODLE, patients from BeSt, NORD-STAR, and RZWN were projected into this pretrained embedding space based on their clinical characteristics at baseline. Each patient was then assigned to the JIP subgroup that had an average clinical profile most closely matching their own, using reference profiles derived from the original LUMC cohort. Beyond baseline, we also projected follow-up visits onto the patient embedding with POODLE. Longitudinal data were collected at months 3, 6, 9, and 12 for BeSt and RZWN and at all time points except nine months for NORD-STAR.

For the NORD-STAR cohort, two adjustments were required. First, longitudinal blood measurements (hemoglobin, leukocytes, thrombocytes, and hematocrit) were not available, so projections were restricted to features shared with the original Leiden embedding (Supplementary Table 1). Second, erythrocyte sedimentation rate (ESR) data were missing for the Danish and Icelandic subcohorts (n = 196), and C-reactive protein was used as a substitute. Prior sensitivity analyses confirmed minimal impact of these adjustments on subgroup assignments (Supplementary Figure S2).[14] Excluding blood measurements had similarly little impact (Supplementary Figure S3), whereas excluding foot-joint measurements negatively impacted assignment (Supplementary Figure S4).

Determining longitudinal stability. Because therapeutic interventions generally reduce joint involvement over time, the JIP distribution was expected to shift during follow-up. To capture this, we monitored changes along the two principal axes of the embedding space: disease extent (oligoarticular vs polyarticular) and anatomic distribution (hand vs foot). To assess longitudinal stability of JIPs, we examined whether patients remained phenotypically closest to the peers they most resembled at baseline. Patient similarity was quantified by calculating cosine distance between any two patients based on their clinical features. Specifically, for each patient at baseline (visit 0), we defined a reference neighborhood using a k-nearest neighbors approach (k = 50) in the JIP-embedding space. This identified the 50 patients with the most similar joint involvement profiles.

At subsequent visits (3, 6, 9, and 12 months), we quantified phenotypic continuity by comparing two distances: the mean cosine distance between a patient and their baseline neighbors (within-subgroup distance) and the mean cosine distance between that patient and all others outside that neighborhood (between-subgroup distance). This ratio of within versus between distance tells us the degree to which patients continue to resemble their original baseline peers phenotypically (compared to the rest of the cohort). We summarized this contrast using a subset coherency (SC) score (a silhouette-like measure[15]) adapted to emphasize local neighborhood structure (specifically, the 50 nearest neighbors) rather than global cluster boundaries, making it less sensitive to boundary effects in the embedding space.

For a given patient i, we defined the following:

- $a(i)$ as the mean cosine distance between patient i and their fixed baseline neighbors (within-neighborhood distance);
- $b(i)$ as the mean cosine distance between patient i and all other patients outside that baseline neighborhood (between-neighborhood distance).

The silhouette-like score was then calculated as follows:

$$s(i) = (b(i) - a(i)) / \max\{a(i), b(i)\}$$

Similar to the silhouette score, the SC score ranges from −1 to 1. A value of 1 indicates maximal phenotypic stability, meaning the patient remains substantially closer to their baseline neighbors than to others. A value near 0 indicates no clear distinction between baseline neighborhood and the rest of the cohort. Negative values indicate divergence, whereby the patient becomes more similar to patients outside their original baseline neighborhood (Figure 1). In practice, silhouette values ≥0.5 are commonly

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

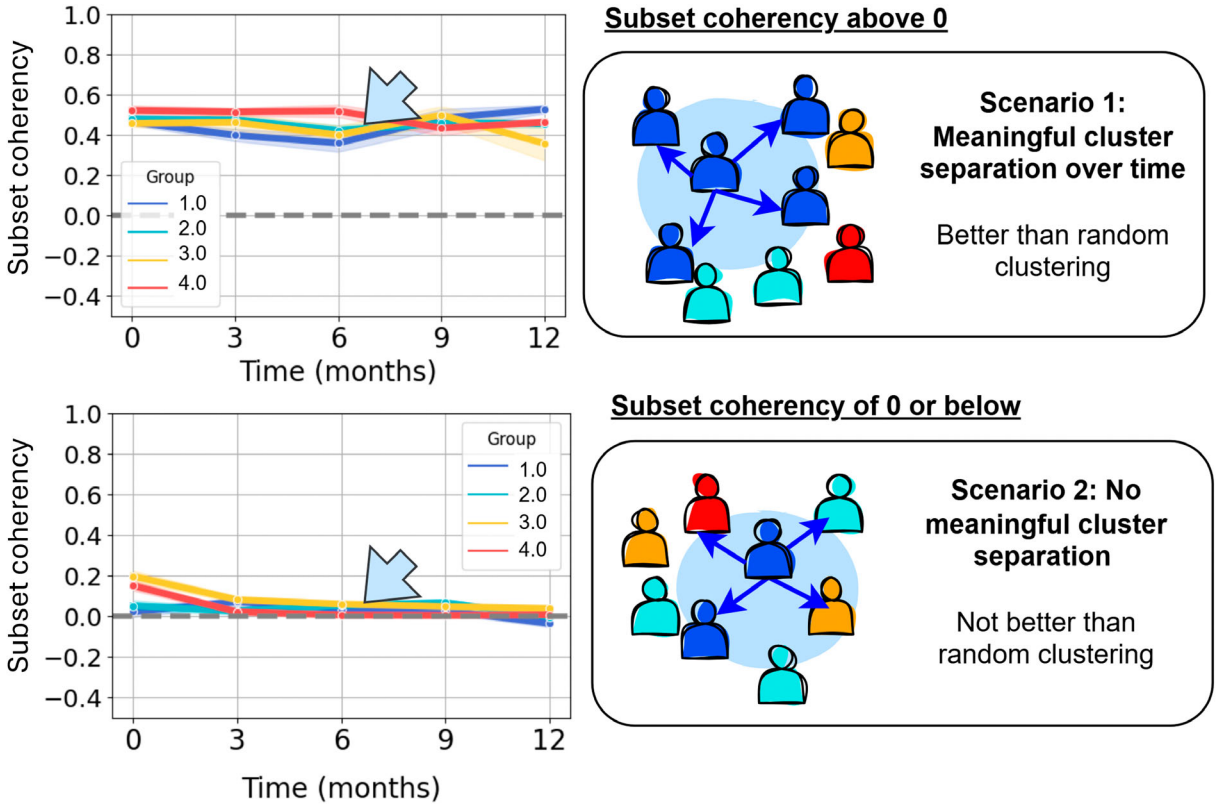

Figure 1. Interpreting longitudinal stability using the subset coherency (SC) score. The 6-month time point is highlighted by a blue arrow, with its corresponding local neighborhood detailed in the right panel, showing a representative patient (blue) surrounded by peers from the same subgroup. Positive SC scores indicate that patients retain similar neighborhoods relative to baseline over time, suggesting meaningful cluster stability, whereas SC scores around or below 0 indicate no meaningful differentiation beyond random clustering. The gray dashed line indicates SC = 0.

interpreted as strong separation and ≥0.25 as moderate separation in clustering contexts[15,16]; however, in longitudinal clinical data subject to treatment-induced homogenization, lower values may still indicate meaningful temporal persistence.

Case vignette: phenotypic stability in diabetes. To illustrate why longitudinal stability is a clinically meaningful property, consider type 1 and type 2 diabetes, two distinct subtypes of the same overarching disease. Viewed cross-sectionally, they can appear strikingly similar: a patient presenting with type 1 diabetes may look indistinguishable from a patient with late-stage type 2 diabetes, particularly when both have progressed to diabetic ketoacidosis (DKA). However, when trajectories are examined over time, the two subtypes clearly diverge. Type 1 follows a rapid, uniform course: patients progress from hyperglycemic symptoms to ketosis within days to weeks. Type 2, by contrast, may remain subclinical for years, with DKA typically only emerging under physiologic stress or advanced insulin deficiency. Crucially, patients within each subtype tend to follow trajectories that resemble one another far more than they resemble those of the other subtype. It is this longitudinal coherence, not the cross-sectional presentation, that validates them as separate phenotypes.

We apply the same logic to our analysis of JIPs in RA. If patients within a given subtype maintain phenotypic similarity over time (that is, if we observe stable subgroup coherence), this supports the interpretation that the subtype reflects a fixed underlying phenotype. Conversely, instability over time would suggest that subgroups capture transient disease stages rather than true distinct entities or that the clinical groups warrant further subdivision.

Change of swollen versus tender joint counts. In addition to the longitudinal stability, we also studied how the disease activity diminished over time. In particular, we studied the change in swollen joint count (SJC) and tender joint count (TJC) and the joint involvement per anatomic region (specifically hand and foot). Regional distribution was assessed by calculating the percentage change in joint count (%ΔJC) relative to baseline for both hand and foot.

To account for the anatomic difference in joint counts among regions, scores were normalized by the total number of evaluable joints in each region based on the DAS66 assessment (DAS28 for feet and DAS32 for hands). This normalization ensures that clinical changes in the feet are represented proportionally to the hands. For instance, resolving pain in three affected joints results in a −10.7% change in the feet (−3/28) compared to a −9.4% change

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

in the hands (–3/32). In this scoring system, 0% represents no change, whereas positive values indicate worsening. Absolute values from which these derivatives were calculated are also reported.

Statistical significance. To confirm that the identified JIP subgroups reflected longitudinally stable patient groupings rather than chance variation, we compared the observed SC score for each cluster against a null distribution generated from 1,000 random clusterings. Empirical *P* values and percentile ranks were derived for each JIP subgroup (median SC) relative to this null distribution at the one-year time point. To compare improvement across anatomic regions, we used a mixed-effects linear regression model with %ΔJC (tenderness or swelling) from baseline (normalized for all evaluable joints per region) as the outcome. The key predictor was the interaction between visit and anatomic region (hand vs foot), which tested whether the rate of improvement differed by location over time. Patient ID was included as a random effect to account for repeated measurements within individuals.

Data availability statement. Data from the BeSt and RZWN cohorts are available upon reasonable request. Access to NORD-STAR data requires submission of a research proposal to the Steering Committee, which will assess its scientific merit and alignment with the study's objectives before granting access. Our code has been published as a repository on GitHub.[17]

RESULTS

Projection on patient embedding. Demographics were broadly similar across cohorts: roughly two-thirds female (64%–69%), mean age at onset between 54 and 61 years, and mean symptom duration around 140 to 165 days (Table 1). However, the BeSt population represented a more severely diseased group compared to the others, with SJC versus TJC of 14 versus 22 in BeSt, 9 versus 12 in NORD-STAR, and considerably lower values of 5 versus 6 in the real-world RZWN cohort. This pattern was also reflected in DAS44 scores, which averaged 4.8 in BeSt, 3.8 in NORD-STAR, and only 2.8 in RZWN. Seropositivity was highest in NORD-STAR (76% rheumatoid factor [RF], 81% anti–cyclic citrullinated peptide antibody [ACPA]), followed by BeSt (65% RF, 61% ACPA), and RZWN had the lowest rates (59.5% RF, 53.3% ACPA). BeSt also had the highest levels of inflammatory activity, with a mean ESR of 37, compared to 28 in NORD-STAR and 13 in RZWN. When comparing patients included in the analysis with those excluded because of missing data, baseline characteristics did not differ meaningfully between the two groups (Supplementary Tables 2–4).

All JIPs recurred across data sets, with each closely resembling its counterpart from the original cohorts (Table 2; Supplementary Figure S5). Specifically, JIP-foot was characterized by younger patients with predominantly foot involvement, JIP-oligo by oligoarticular disease with few affected joints, JIP-hand by older patients with predominantly hand involvement, and JIP-poly by severe, widespread disease affecting many joints, including the hands, feet, and large joints. However, the distribution of patients across the two embedding axes—the degree of involvement and anatomic region—differed among the three cohorts (Supplementary Figure S6).

In the BeSt cohort, most patients clustered near the JIP-poly subtype in the patient embedding space, likely reflecting their greater disease severity compared to the original Leiden cohort on which the clinical JIP embedding was based. BeSt patients had higher median ESR (37 vs 25 mm/hr), SJC (13 vs 7), and TJC (13 vs 10) and DAS scores (4.3 vs 3.5). Consequently, most

Table 1. Baseline characteristics of the BeSt, NORD-STAR, and RZWN cohorts, compared to the Leiden cohort as reference*

Characteristic	Original cohort (reference)	BeSt	NORD-STAR	RZWN
N	1,387	445	630	210
Female ɣ, n (%)	894 (64.5)	300 (67.4)	439 (69.7)	145 (69.0)
Age ɣ (SD), y	60 (45–75)	54 (44–64)	56 (46–66)	61 (47–75)
RF ɣ, n (%)	722 (52.1)	291 (65.4)	479 (76.0)	125 (59.5)
ACPA ɣ, n (%)	676 (48.7)	271 (60.9)	516 (81.9)	112 (53.3)
ESR ɣ (IQR), mm/hr	25 (11–45)	37 (20–55)	28 (18–43)[a]	13 (7–24)
DAS (3) (IQR)	3.5 (2.7–4.3)	4.8 (4.1–5.9)	3.8 (3.1–4.7)	2.8 (1.9–3.9)
SJC (IQR)	7 (4–12)	14 (10–20)	9 (5–15)	5 (1–13)
TJC (IQR)	10 (5–17)	22 (14–32)	12 (8–21)	6 (2–14)
DAS28 (3) (IQR)	5.2 (4.3–6.2)	6.5 (5.8–7.3)	5.6 (4.7–6.4)	4.3 (3.4–5.5)
Symptom duration (IQR), d	155 (62–512)	165 (96–376)	141 (88–215)	–

* To identify joint involvement pattern clusters in variables, ɣ was used. ACPA, anti–cyclic citrullinated peptide antibody; BeST, Behandel-Strategieën; DAS28 (3), three-component Disease Activity Score in 28 joints; ESR, erythrocyte sedimentation rate; IQR, interquartile range; NORD-STAR, Nordic Rheumatic Diseases Strategy Trials and Registries; RF, rheumatoid factor; RZWN, Reumazorg Zuidwest Nederland; SJC, swollen joint count; TJC, tender joint count.
[a] ESR was calculated for 530 patients with available measurements (Danish patients were excluded because we only had C-reactive protein measured in set B).

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. Cluster characteristics for the BeSt, NORD-STAR, and RZWN*

	BeSt (n = 445)				NORD-STAR (n = 630)				RZWN (n = 210)			
Characteristic	JIP-foot	JIP-oligo	JIP-hand	JIP-poly	JIP-foot	JIP-oligo	JIP-hand	JIP-poly	JIP-foot	JIP-oligo	JIP-hand	JIP-poly
n	60	67	129	189	190	177	99	164	38	121	39	12
Female ɣ, n (%)	42 (70.0)	43 (64.2)	83 (64.3)	132 (69.8)	137 (72.1)	126 (71.2)	66 (66.7)	110 (67.1)	27 (71.1)	87 (71.9)	21 (53.8)	10 (83.3)
Age ɣ (SD), y	50 (37–64)	51 (40–63)	56 (42–70)	54 (41–68)	50 (36–64)	56 (44–68)	64 (54–74)	54 (38–70)	65 (54–76)	59 (45–73)	67 (55–79)	52 (37–67)
RF ɣ, n (%)	39 (65)	41 (61.2)	89 (69.0)	122 (64.6)	156 (82.1)	138 (78.0)	70 (70.7)	115 (70.1)	22 (57.9)	71 (58.7)	23 (59.0)	9 (75.0)
ACPA ɣ, n (%)	39 (65)	43 (64.2)	81 (62.8)	108 (57.1)	170 (89.5)	162 (91.5)	66 (66.7)	118 (72.0)	18 (47.4)	70 (57.9)	17 (43.6)	7 (58.3)
ESR ɣ (IQR), mm/hr	28 (14–45)	31 (20–42)	39 (24–60)	40 (20–59)	23 (12–34)	27 (18–42)	34 (22–49)	32 (20–49)	17 (8–25)	12 (7–20)	13 (7–30)	28 (15–34)
DAS (IQR)	4.1 (3.7–4.7)	3.5 (3.2–3.9)	4.4 (3.9–4.9)	5.1 (4.5–5.7)	3.6 (3.1–4.2)	2.8 (2.4–3.1)	3.9 (3.6–4.4)	4.8 (4.1–5.4)	3.3 (2.9–3.9)	2.0 (1.5–2.4)	3.4 (3.0–4.0)	4.5 (3.9–5.1)
SJC (IQR)	12 (10–17)	9 (7–11)	13 (10–18)	18 (12–24)	9 (6–12)	4 (2–7)	12 (8–14)	18 (12–24)	9 (8–12)	2 (1–5)	12 (8–14)	16 (14–23)
TJC (IQR)	20 (14–25)	11 (7–16)	18 (13–24)	32 (23–45)	13 (9–19)	6 (3–9)	13 (9–18)	25 (17–32)	10 (6–14)	2 (1–4)	9 (5–13)	17 (10–27)
DAS28 (IQR)	5.2 (4.4–5.9)	5.1 (4.7–5.4)	5.9 (5.4–6.4)	6.0 (5.3–6.6)	4.8 (4.1–5.4)	4.3 (3.8–4.9)	5.6 (5.2–6.1)	5.9 (5.1–6.6)	4.1 (3.4–4.8)	2.9 (2.2–3.7)	5.0 (4.0–5.4)	5.1 (4.5–5.9)
Symptom duration (IQR), d	166 (107–275)	247 (117–481)	160 (89–295)	159 (94–386)	119 (81–196)	142 (91–224)	152 (85–201)	149 (95–220)	n.a.	n.a.	n.a.	n.a.

* To identify joint involvement pattern (JIP) clusters in variables, ɣ was used. ACPA, anti–cyclic citrullinated peptide antibody; BeST, Behandel-Strategieën; DAS28, Disease Activity Score in 28 joints; ESR, erythrocyte sedimentation rate; IQR, interquartile range; JIP-foot, JIP in foot arthritis; JIP-hand, JIP in hand arthritis; JIP-oligo, JIP in oligoarthritis; JIP-poly, JIP in polyarthritis; n.a., not available; NORD-STAR, Nordic Rheumatic Diseases Strategy Trials and Registries; RF, rheumatoid factor; RZWN, Reumazorg Zuidwest Nederland; SJC, swollen joint count; TJC, tender joint count.

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

BeSt patients were assigned to the JIP-poly subset (JIP-foot, 60; JIP-oligo, 67; JIP-hand, 129; JIP-poly, 189; Table 2). Symptom duration was comparable across clusters (156–159 days), except for JIP-oligo, which had the longest duration (247 days).

Despite NORD-STAR also being a trial cohort, the patient population more closely resembled the original Leiden cohort. They exhibited slightly elevated markers of inflammation (ESR 28 vs 25 mm/hr) and joint involvement (SJC, 9 vs 7; TJC, 12 vs 10) compared to the original cohort, but they had nearly identical DAS values (3.8 vs 3.5). This corresponded to a different distribution of clusters (JIP-foot, 190; JIP-oligo, 177; JIP-hand, 99; JIP-poly, 164; Table 2). Symptom duration was comparable among the clusters (119–152 days). Unlike the two trials, RZWN patients predominantly had mild disease, clustering near the JIP-oligo pole, with lower markers of inflammation, joint involvement (ESR 16 vs 25 mm/hr; SJC, 5 vs 7; TJC, 6 vs 10) and DAS scores (2.8 vs 3.5) compared to the original Leiden cohort. Consequently, most samples belonged to the oligoarticular phenotype (JIP-foot, 38; JIP-oligo, 121; JIP-hand, 39; JIP-poly, 12; Table 2).

Longitudinal changes in joint involvement. After treatment initiation, joint counts and DAS scores generally decreased across all cohorts (Figure 2; Supplementary Figures S7 and S8). In the RZWN cohort, this reduction was primarily observed in SJCs. These trends likely reflect the success of treat-to-target strategies, with the majority of patients transitioning toward a mild disease state characterized by minimal swelling and tenderness.

Temporal analysis of cluster consistency (Figure 3) revealed that clusters were most coherent at baseline (mean SC = 0.70–0.80), indicating that patients initially aligned more closely with their baseline phenotype neighbors than with other groups. Following baseline, cluster consistency declined in the BeSt and NORD-STAR cohorts, potentially due to the substantial reduction in the total number of affected joints. Despite this decline, clusters generally maintained higher coherence over one year than would be expected by chance, as demonstrated by the permutation analysis (mean SC = 0.25 in BeSt, with P values ranging from 0.018 to 0.038, and SC = 0.26 in NORD-STAR, with P = 0.004–0.017). With the exception of the JIP-poly group in BeSt (SC = 0.06; P = 0.280) and the JIP-oligo group in NORD-STAR (SC = 0.04; P = 0.278)—both of which lost coherence after six months—patients remained more similar to their baseline neighbors after one year than to patients with different initial joint patterns.

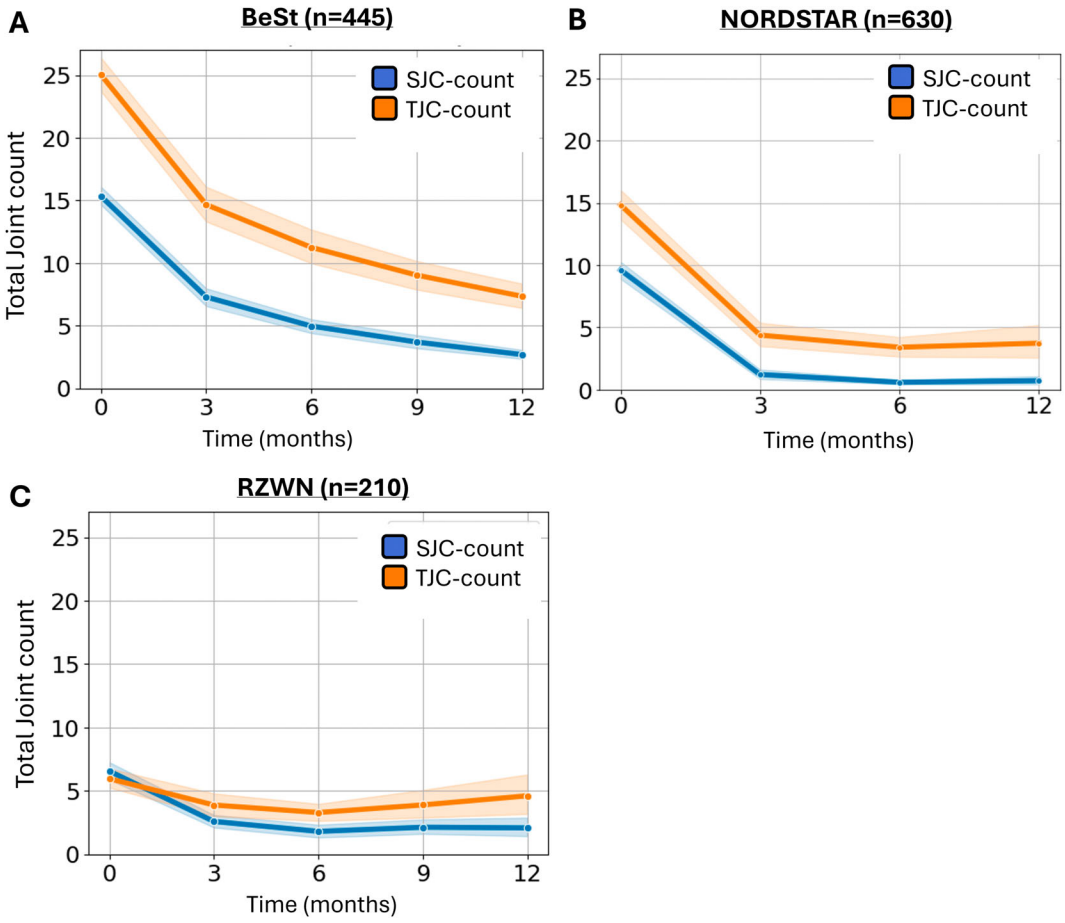

Figure 2. Mean tender joint count (TJC) and swollen joint count (SJC) over one year in (A) Behandel-Strategieën (BeSt), (B) Nordic Rheumatic Diseases Strategy Trials and Registries (NORD-STAR), and (C) Reumazorg Zuidwest Nederland (RZWN) cohorts.

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

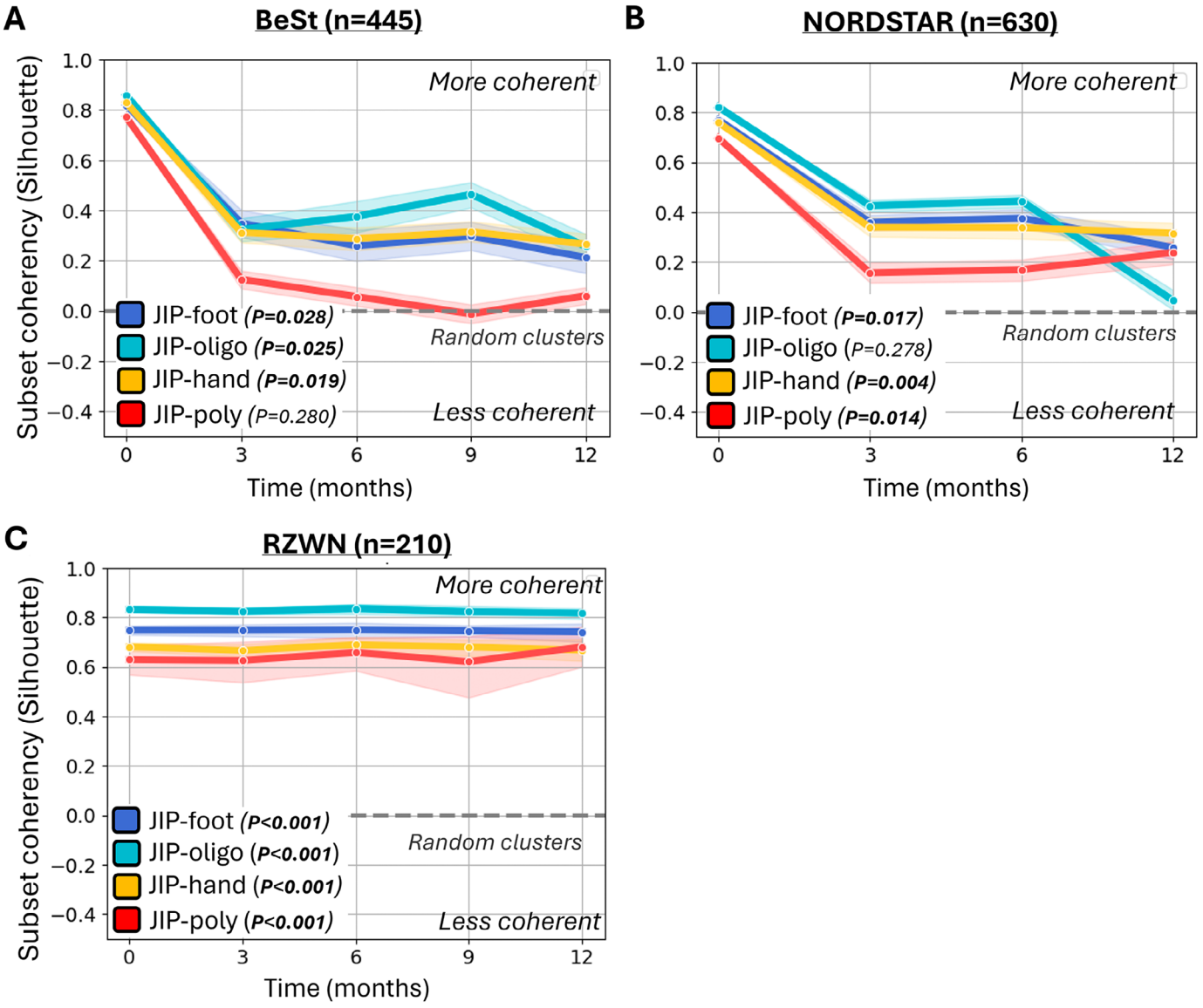

Figure 3. Longitudinal stability of joint involvement patterns in (A) Behandel-Strategieën (BeSt), (B) Nordic Rheumatic Diseases Strategy Trials and Registries (NORD-STAR), and (C) Reumazorg Zuidwest Nederland (RZWN) cohorts. Each panel illustrates the degree of phenotypic similarity across one year among patients sharing the same baseline joint involvement pattern (JIP) compared to those from different baseline subsets. The metric is based on a silhouette-like score (a cluster validity index), which ranges from −1.0 to 1.0. The dashed line at 0.0 indicates no coherence (no discernible subgroups); values approximating 1.0 indicate great coherence, and values approximating −1.0 indicate great incoherence (drift toward different subgroups). JIP-foot, JIP in foot arthritis; JIP-hand, JIP in hand arthritis; JIP-oligo, JIP in oligoarthritis; JIP-poly, JIP in polyarthritis.

Comparing within-subgroup to between-subgroup similarity yielded consistent findings (Supplementary Table 5; Supplementary Figure S9). In BeSt, patients in the JIP-hand subgroup retained 26.2% (95% CI 22.0–30.5) greater overlap in clinical features with their baseline neighbors than with the broader cohort, followed closely by JIP-oligo (25.5%; 95% CI 20.5–30.5) and JIP-foot (20.3%; 95% CI 13.5–27.0). In NORD-STAR, the JIP-hand subgroup again showed the strongest coherence, with members remaining 31.5% (95% CI 26.9–36.0) more phenotypically similar to each other than to the rest of the cohort, followed by JIP-foot (24.5%; 95% CI 19.4–29.7) and JIP-poly (22.6%; 95% CI 17.5–27.6).

Interestingly, we observe a difference between trial data and real-world data (RZWN). In the trial data, group coherence gradually decreases over time (mean SC 0.76 and 0.81 vs 0.25 and 0.26), likely reflecting the effects of intensive therapy or patient selection. As disease activity diminishes, phenotypic differences among patients become less pronounced. In contrast, the subgroups remained considerably more longitudinally stable in RZWN (mean SC 0.75; all $P < 0.001$). Among these, JIP-oligo showed the highest cluster distinctiveness, remaining 78.1% (95% CI 75.8–80.4) more phenotypically similar to its baseline neighbors than to the rest of the cohort, followed by JIP-foot

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(74.1%; 95% CI 69.7–78.6), JIP-poly (65.6%; 95% CI 43.8–82.8), and JIP-hand (63.4%; 95% CI 57.7–69.2).

Hand and foot asymmetry. In contrast to the overall reduction in total joint counts (Figure 2), which primarily characterizes the first dimension of the JIP embedding (eg, polyarticular vs oligoarticular or severe vs mild disease), the second dimension (representing hand–foot differentiation) alluded to an asymmetric pattern of improvement depending on the anatomic region. Specifically, in both the BeSt and NORD-STAR cohorts, the proportional reduction in affected foot joints lagged behind the reduction observed in hand joints. This was observed for both tender and swollen joints (Figure 4; Supplementary Figures S10 and S11). However, the site-specific disparity did not appear to translate to the real-world RZWN cohort, wherein comparable rates of reduction were observed across both anatomic dimensions.

DISCUSSION

Several RA subsets have been proposed, and if these reflect truly distinct phenotypes, they would imply unique underlying pathologic mechanisms and potentially require tailored therapeutic approaches. However, how to distinguish disease subsets from disease states remains a central unresolved question in the field.[5] Our study introduces a novel method to study this longitudinal stability, even in a treat-to-target setting in which therapy is explicitly designed to drive patients toward a similar mild expression of disease. The JIPs were no exception to this dynamic. As treat-to-target therapy took effect, joint counts declined and phenotypic differences among subtypes progressively narrowed. Nonetheless, patients with similar baseline joint involvement continued to cluster together throughout follow-up.

In our trial populations, patients remained on average 20% to 30% more phenotypically similar to their baseline neighbors than to the broader cohort, a pattern even more pronounced in the

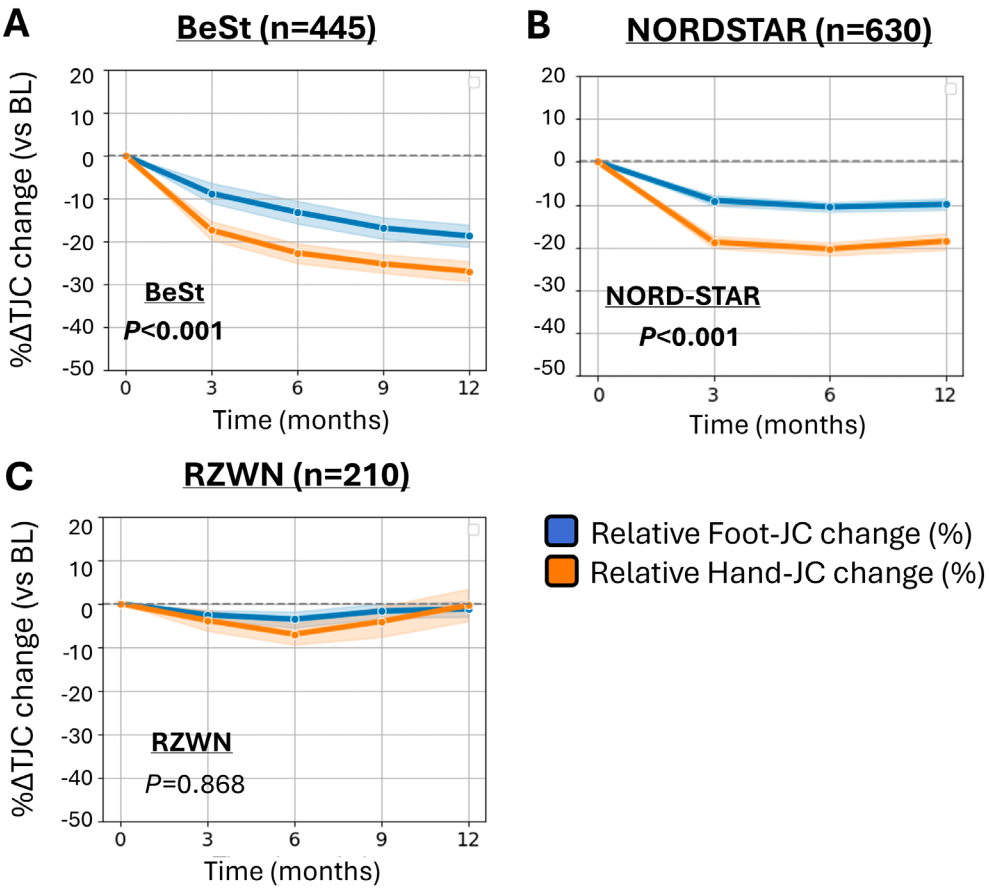

Figure 4. Relative tender joint count (TJC) changes in hand and feet across the (A) Behandel-Strategieën (BeSt), (B) Nordic Rheumatic Diseases Strategy Trials and Registries (NORD-STAR), and (C) Reumazorg Zuidwest Nederland (RZWN) cohorts. Values are expressed as percentage change in TJCs (%ΔTJC) from baseline (BL), normalized to the total number of evaluable hand and foot joints per assessment method. Negative values denote clinical improvement, whereas positive values indicate disease progression. Statistical significance was inferred using a mixed-effects linear regression model, with individual patients treated as random effects to account for repeated measures. Reported *P* values correspond to the interaction term between visit ID and anatomic region (hand vs foot), evaluating whether the rate of improvement significantly differed by location.

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

real-world setting (60%–80%). This suggests that patients retain meaningful phenotypic resemblance to their original cluster neighbors even as disease activity improves, pointing to a shared underlying pathology. This supports the rationale for precision medicine, in which subgroup-specific interventions rely on the assumption that a patient's phenotype reflects stable underlying pathology rather than a temporary phase. If these patterns merely reflected disease stages, we would expect the SC to disappear next visit; patients would redistribute across other subgroups irrespective of their baseline assignment. Instead, patients maintained significantly higher coherence to their baseline neighbors than expected by chance (confirmed by 1,000 random permutations), even following intensive treat-to-target therapy.

A counterargument, however, is that these clusters could still represent sequential stages along a shared disease trajectory, with patients entering cohorts at different points. For example, one could hypothesize a preclinical progression from JIP-oligo to JIP-hand/foot and ultimately to JIP-poly, such that the identified clusters reflect positions along this trajectory. Under this model, apparent stability would arise because patients move gradually along the same continuum over time, maintaining phenotypic similarity as they transition between adjacent stages.

If the clusters indeed represented different stages, we would expect systematic differences in symptom duration, with JIP-oligo showing the shortest. This was not observed; symptom duration did not differ among the JIPs. Notably, in BeSt, JIP-oligo (which would be deemed an early stage in this hypothetical situation) showed the longest symptom duration. Moreover, the previously reported tendency for inflammation to recur in the same joints over time[18] adds further evidence toward these joint patterns being stable phenotypes.

Although our subgroups demonstrated phenotypic continuity across the cohorts, notable differences emerged between the real-world cohort and the trial cohorts. In the trial cohorts, subgroup coherency declined following treatment initiation, consistent with the intensive treat-to-target strategies employed in clinical trials. This was not observed in RZWN, possibly reflecting less intensive treatment in routine care, lower baseline disease severity (because patients were not required to meet 1987 or 2010 classification criteria), or both. Prospective studies directly comparing these paradigms would be needed to disentangle the contribution of each factor separately. To facilitate the evaluation of longitudinal robustness in clinical subgroups, we have made our method publicly available on GitHub.[17] For direct application of JIP classification, we provide a user-friendly web tool that requires no programming expertise.[19]

Beyond subgroup stability, a closer inspection of the changes at the anatomic regions characteristic of these subgroups revealed a recurring pattern: both BeSt and NORD-STAR, but not RZWN, showed asymmetric improvements in tenderness and swelling with foot joints lagging behind hand joints. Recent evidence supports this asymmetric pattern of joint response. Wechalekar et al[20] reported persistent foot-joint synovitis in 25% to 36% of patients otherwise meeting remission criteria, and Ciurea et al[21] showed less sustained reductions in foot-joint tenderness and swelling compared to hand joints over 12 months of biologic DMARD therapy.

A biologic basis for this joint-specific variation may lie in the transcriptomic diversity identified by Ai et al,[22] who characterized distinct synovial fibroblast populations across anatomic joint locations. Additionally, some differences might be explained by noninflammatory processes; specifically, site-specific fibroblast activation may drive increased innervation and pain that are disproportionate to the actual degree of synovitis.[23] Supporting this, Bai et al[24] demonstrated in mouse models that certain fibroblast subpopulations promote nerve growth and pain sensitivity.

Nonetheless, further research into joint-specific pathogenic mechanisms is warranted. The lack of hand–foot asymmetry in RZWN, for instance, may reflect early treatment initiation in this real-world setting, potentially resolving foot involvement before it fully manifests. This hand–foot differentiation highlights a clinical implication: foot joints should be assessed alongside the hands. Yet, many clinical trials rely on composite scores (eg, DAS28, Clinical Disease Activity Index, and Simplified Disease Activity Index) that omit foot joints, entirely risking systematic underrepresentation of a clinically prevalent[25] and mechanistically distinct dimension—one that may also represent a treatment target. Supporting this possibility, a previous observational study across the BeSt and NORD-STAR cohorts[14] found that hand-dominant JIP was associated with the best treatment response, whereas JIP-poly was associated with the worst, consistent with the original findings.[4]

Several limitations of this study should be considered. First, subset coherence at one year was modest, particularly in the trial cohorts (mean 0.25). This suggests that the JIP markers do not fully capture the variation observed over time, such as differences between treatment responders and nonresponders. On a related note, the coherence of these subsets may decline over longer follow-up, although the observed plateau in stability curves suggests some persistence. Second, we did not account for the effect of different treatment regimens across the three cohorts because stratifying by treatment would have left insufficient sample sizes—particularly when examined across all clusters.

Third, the generalizability of our subset coherence may be restrained by patient selection and related biases. Even across our cohorts, the criteria used to select patients with RA differed. Furthermore, differences in health care organization, recruitment practices, and clinical care may limit the extent to which the observed longitudinal stability translates to other settings.[26–28] Although we assessed concordance of the JIP clusters across both clinical trial and real-world populations, all cohorts were derived from Northern European health care settings.

Similarly, missing follow-up data may be biased toward certain patient characteristics or by physician-related factors. Although we previously found no evidence of physician-related

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

bias affecting the JIPs in the original cohort[4] and the baseline characteristics did not differ substantially between patients included and those excluded because of missing data, we cannot exclude this possibility entirely. Future studies in more diverse geographic and clinical settings are needed to assess the generalizability of our findings.

In conclusion, we developed a novel method to assess longitudinal stability and showed that JIP subgroups remain coherent. Despite effective therapy reducing disease activity, patients remained more similar to those with the same baseline joint pattern than to others. Together with the absence of baseline differences in symptom duration, this persistence suggests that JIP classification may capture stable phenotypes rather than simply representing different stages. This is an important principle to establish in light of the precision medicine paradigm: clinically meaningful, stable phenotypes exist in RA, potentially reflecting distinct underlying mechanisms. Future studies are needed to elucidate the biologic mechanisms underlying these phenotypes and to determine whether they can inform the selection of more tailored therapeutic strategies. To facilitate further research, our method for studying longitudinal stability is now publicly available.

ACKNOWLEDGMENTS

The authors used Claude.ai (Anthropic) to improve readability and check spelling and grammar. After using this tool, the content was subsequently reviewed and edited by the authors, who take full responsibility for the final version of this article.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Knevel confirms that all authors have provided the final approval of the version to be published and takes responsibility for the affirmations regarding article submission (eg, not under consideration by another journal), the integrity of the data presented, and the statements regarding compliance with institutional review board/Declaration of Helsinki requirements.

REFERENCES

1. Firestein GS. The disease formerly known as rheumatoid arthritis. Arthritis Res Ther 2014;16(3):114. doi:https://doi.org/10.1186/ar4593
2. Lewis MJ, Barnes MR, Blighe K, et al. Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. Cell Rep 2019;28(9):2455–2470.e5. doi:https://doi.org/10.1016/j.celrep.2019.07.091
3. Zhang F, Jonsson AH, Nathan A, et al. Cellular deconstruction of inflamed synovium defines diverse inflammatory phenotypes in rheumatoid arthritis. bioRxiv. Published online January 1, 2022:2022.02.25.481990. doi:https://doi.org/10.1101/2022.02.25.481990
4. Maarseveen TD, Maurits MP, Coletto LA, et al. Location and amount of joint involvement differentiates rheumatoid arthritis into different clinical subsets. NPJ Digit Med 2025;8(1):623. doi:https://doi.org/10.1038/s41746-025-01997-1
5. Triaille C, Durez P, Natalucci F, et al. The relationship between clinical disease activity, synovial inflammatory profile, and treatment response in rheumatoid arthritis. Lancet Rheumatol 2026;8:e66–e73. doi:https://doi.org/10.1016/S2665-9913(25)00252-8
6. Goekoop-Ruiterman YPM, de Vries-Bouwstra JK, Allaart CF, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis Rheum 2005;52(11):3381–3390. doi:https://doi.org/10.1002/art.21405
7. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31(3):315–324. doi:https://doi.org/10.1002/art.1780310302
8. Østergaard M, van Vollenhoven RF, Rudin A, et al; NORD-STAR study group. Certolizumab pegol, abatacept, tocilizumab or active conventional treatment in early rheumatoid arthritis: 48-week clinical and radiographic results of the investigator-initiated randomised controlled NORD-STAR trial. Ann Rheum Dis 2023;82(10):1286–1295. doi:https://doi.org/10.1136/ard-2023-224116
9. Hetland ML, Haavardsholm EA, Rudin A, et al; NORD-STAR study group. Active conventional treatment and three different biological treatments in early rheumatoid arthritis: phase IV investigator initiated, randomised, observer blinded clinical trial. BMJ. 2020;371:m4328. doi:https://doi.org/10.1136/bmj.m4328
10. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569–2581. doi:https://doi.org/10.1002/art.27584
11. Maarseveen TD, Selani D, Steinz N, et al. Work smarter, not harder: achieve expert-level diagnosis extraction from medical records with optimal prompting of large language models. Ann Rheum Dis 2026;85:778–780. doi:https://doi.org/10.1016/j.ard.2025.11.001
12. Maarseveen TD. POODLE: Projecting Observations on a Deep Learned Embedding. GitHub. August 8, 2022. Accessed March 19, 2026. https://github.com/levrex/Poodle
13. Onken A, Yague J, Sakata S. Deep multimodal autoencoders for identifying latent representations of spike counts and local field potentials. Poster session presented at: Bernstein Conference; September 25–28, 2018; Berlin, Germany. doi:https://doi.org/10.12751/nncn.bc2018.0223
14. Nagafuchi Y, Maarseveen TD, Lend K, et al. Hand-dominant joint involvement pattern associates with favourable, and polyarthritis with unfavourable, treatment response to both csDMARDs and bDMARDs in early rheumatoid arthritis: a combined analysis of NORD-STAR and BeSt trials. Ann Rheum Dis 2026;85:1218–1228. doi:https://doi.org/10.1016/j.ard.2026.02.005
15. Dalmaijer ES, Nord CL, Astle DE. Statistical power for cluster analysis. BMC Bioinformatics 2022;23(1):205. doi:https://doi.org/10.1186/s12859-022-04675-1
16. Struyf A, Hubert M, Rousseeuw PJ. Integrating robust clustering techniques in S-PLUS. Comput Stat Data Anal 1997;26(1):17–37. doi:https://doi.org/10.1016/S0167-9473(97)00020-0
17. Heckert SL, Bergstra SA, Matthijssen XME, et al. Joint inflammation tends to recur in the same joints during the rheumatoid arthritis disease course. Ann Rheum Dis 2022;81(2):169–174. doi:https://doi.org/10.1136/annrheumdis-2021-220882
18. Maarseveen TD. Temporal analysis subset. Github. December 18, 2022. Accessed April 16, 2026. https://github.com/levrex/Temporal_analysis_subsets
19. Knevel Lab. Rheumalyze. Github. Accessed April 14, 2026. https://github.com/Knevel-Lab/Rheumalyze
20. Wechalekar MD, Lester S, Proudman SM, et al. Active foot synovitis in patients with rheumatoid arthritis: applying clinical criteria for disease activity and remission may result in underestimation of foot joint

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

involvement. Arthritis Rheum 2012;64(5):1316–1322. doi:https://doi.org/10.1002/art.33506

21. Ciurea A, Distler O, Kwok K, et al. Joint-level responses to tofacitinib and methotrexate: a post hoc analysis of data from ORAL Start. Arthritis Res Ther 2023;25(1):185. doi:https://doi.org/10.1186/s13075-023-03144-1
22. Ai R, Hammaker D, Boyle DL, et al. Joint-specific DNA methylation and transcriptome signatures in rheumatoid arthritis identify distinct pathogenic processes. Nat Commun 2016;7(1):11849. doi:https://doi.org/10.1038/ncomms11849
23. Das D, Choy E. Non-inflammatory pain in inflammatory arthritis. Rheumatology (Oxford) 2023;62(7):2360–2365. doi:https://doi.org/10.1093/rheumatology/keac671
24. Bai Z, Bartelo N, Aslam M, et al; Accelerating Medicines Partnership RA/SLE Network. Synovial fibroblast gene expression is associated with sensory nerve growth and pain in rheumatoid arthritis. Sci Transl Med 2024;16(742):eadk3506. doi:https://doi.org/10.1126/scitranslmed.adk3506
25. Lee SW, Kim SY, Chang SH. Prevalence of feet and ankle arthritis and their impact on clinical indices in patients with rheumatoid arthritis: a cross-sectional study. BMC Musculoskelet Disord 2019;20(1):420. doi:https://doi.org/10.1186/s12891-019-2773-z
26. June RR, Aggarwal R. The use and abuse of diagnostic/classification criteria. Best Pract Res Clin Rheumatol 2014;28(6):921–934. doi:https://doi.org/10.1016/j.berh.2015.04.004
27. Aggarwal R, Ringold S, Khanna D, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken) 2015;67(7):891–897. doi:https://doi.org/10.1002/acr.22583
28. Gomon G, Raffray M, Betancort Rodríguez C, et al. Referral pathways and duration of care for musculoskeletal complaints across Europe: an analysis of primary and secondary care. Ann Rheum Dis 2026;0:S0003-4967(26)00289-X. doi:https://doi.org/10.1016/j.ard.2026.05.011

21514658, 0, Downloaded from https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80174, Wiley Online Library on [18/11/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License