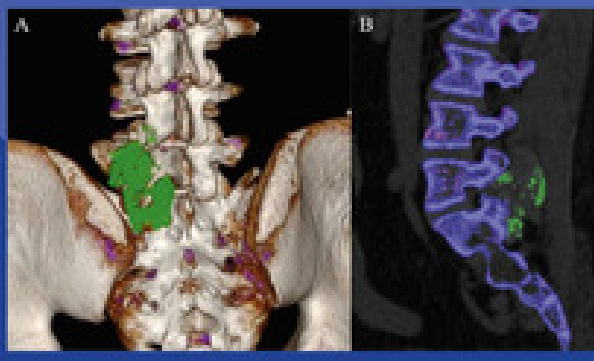
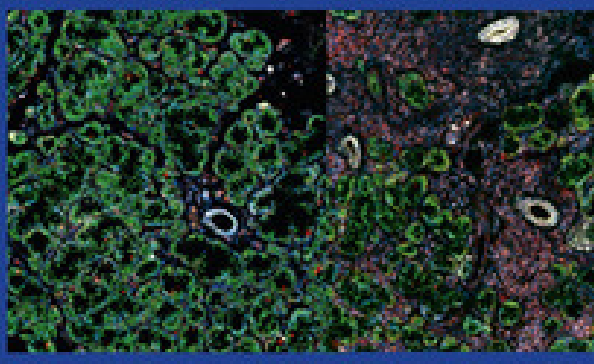
APRIL 2025
VOLUME 84
ISSUE
4

ANNALS OF THE RHEUMATIC DISEASES

THE EULAR JOURNAL





ELSEVIER

eular
EUROPEAN ALLIANCE OF ASSOCIATIONS FOR RHEUMATOLOGY

The home of innovative rheumatology research

Ann Rheum Dis 84 (2025) 519–520

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Editorial

Methotrexate osteopathy: an increasingly recognised condition manageable only through methotrexate discontinuation

There is increasing evidence that low-dose methotrexate (MTX) for the treatment of rheumatic musculoskeletal diseases can lead to a unique type of atraumatic insufficiency fracture of the lower limb, termed MTX osteopathy. A study now published in this issue of *Annals of the Rheumatic Diseases* [1] sheds light on the importance of stopping MTX to manage this rare skeletal side effect.

A RARE CONDITION FIRST DESCRIBED MORE THAN 50 YEARS AGO

Fractures secondary to long-term MTX therapy were first described in 1970 in children with acute lymphoblastic leukaemia in remission [2]. These children presented with a triad of symptoms consisting of bone pain, osteoporosis, and fractures. Radiographs of the knee joints showed metaphyseal osteoporosis, contrasting with abnormally dense zones of provisional calcification. Importantly, this report already described that discontinuation of MTX was followed by skeletal remineralisation of the lower extremity. More than 20 years later, MTX osteopathy was first described in two patients with rheumatic musculoskeletal diseases who were treated with low-dose MTX for rheumatoid arthritis or psoriasis [3]. These patients had atraumatic insufficiency fractures of the proximal and distal tibiae with recurrent symptoms that resolved after discontinuation of MTX.

CURRENT EVIDENCE

The current study by Hauser et al [1] provides a comparatively large case series of 33 patients with insufficiency fractures associated with MTX therapy and confirms their key demographic and disease-specific characteristics. These are almost exclusively women, mostly in their mid-60s, with rheumatoid arthritis or other rheumatic musculoskeletal diseases who have been treated with MTX for several years. As detailed in another previously published case series of 34 patients [4] and a systematic review of 80 patients [5], the most commonly affected skeletal sites are the distal tibia, calcaneus, and proximal tibia, with a pathognomonic fracture morphology on magnetic resonance imaging described as band-like and along the growth plate appearance. Importantly, although most patients have osteoporosis confirmed by dual-energy X-ray absorptiometry, only a few are concomitantly treated with glucocorticoids, providing additional evidence that the observed fractures may be mechanistically distinct from other fractures observed in patients with rheumatic diseases. In sum, the collective evidence has shown that low-dose MTX therapy can lead to atraumatic stress fractures (or more accurately, insufficiency fractures according to current nomenclature [6]) of the lower extremity, which may mimic arthritis.

THE IMPORTANCE OF STOPPING MTX AND FURTHER TREATMENT CONSIDERATIONS

The main novelty of the study by Hauser et al [1] lies in the valuable observation that almost all patients who continued MTX (95%, 20 of 21 patients) suffered further fractures, including insufficiency fractures as well as typical fragility fractures, whereas further fractures were observed in only 36% (4 of 11) of patients in who discontinued MTX. As insufficiency fractures in MTX osteopathy often occur at multiple skeletal sites [4] and bilaterally [7], the prevention of additional fractures is often crucial to maintain some degree of weight bearing and mobilisation. In addition, Hauser et al [1] found a significant improvement in pain levels and weight bearing tolerance, indirect indicators of fracture healing, in the affected skeletal region in patients who discontinued MTX. These findings are supported by a case report of a woman with systemic lupus erythematosus and recurrent bilateral insufficiency fractures of the calcaneus due to long-term MTX use, in whom fracture healing and successful remobilisation were achieved only by MTX discontinuation [8]. As most patients with MTX osteopathy have osteoporosis, antiosteoporotic treatment is virtually always indicated. At present, it remains to be determined which osteoporosis drug is most effective in this specific patient cohort; however, osteoanabolic therapy appears to be the most promising due to its high potency. Respective treatment concepts may favour the recombinant human parathyroid hormone analogue teriparatide (optionally combined with the receptor activator of nuclear factor kappa-B ligand antibody denosumab [9]) or the sclerostin antibody romosozumab [10]. Another important question for rheumatologists is whether MTX needs to be replaced by another disease-modifying antirheumatic drug and, if so, which one. However, there is no one-size-fits-all solution, as disease activity often varies considerably in individual patients, and the diversity of rheumatic diseases, which originally

https://doi.org/10.1016/j.ard.2025.01.041
Received 6 January 2025; Revised 23 January 2025; Accepted 26 January 2025

justified the use of MTX, requires individualised approaches in the era of biologic therapies. At best, some of these therapies even have a positive systemic effect on bone outcomes [11], but this assumption requires further research.

OUTLOOK

The study by Hauser et al [1] adds to the growing body of evidence supporting the existence of MTX osteopathy, but pathomechanistic explanations for this phenomenon are still lacking. Only few studies have previously reported indications of impaired bone remodelling by MTX, namely reduced bone formation *in vitro* [12] and impaired bone mechanical properties and increased bone resorption in animal models [13]. Clinically, however, it is striking that MTX is used in hundreds of thousands of people worldwide without causing insufficiency fractures. This suggests that, unlike the adverse effects of glucocorticoids, which affect the vast majority of people treated [14], only a few people are susceptible to MTX osteopathy. This may imply, for example, that genetic variants increase susceptibility to MTX-associated insufficiency fractures, as has previously been reported for susceptibility to bisphosphonate-related atypical femoral fractures [15]. Finally, in addition to conducting basic research on the pathomechanism of MTX osteopathy, it seems most important to conduct prospective comparative and randomised clinical trials to investigate both future fracture risk and fracture healing with MTX continuation versus MTX discontinuation and to compare the effects of different antiosteoporotic drugs.

Competing interests

The author declares no competing interests.

Funding

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

Patient consent for publication

Not applicable.

Ethics approval

Not applicable.

Provenance and peer review

Not commissioned; externally peer reviewed.

Orcid

Tim Rolvien: http://orcid.org/0000-0003-1058-1307

Tim Rolvien *
Department of Trauma and Orthopaedic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
***Correspondence to** Dr Tim Rolvien, Department of Trauma and Orthopaedic Surgery, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
E-mail address: t.rolvien@uke.de

REFERENCES

[1] Hauser B, Merriman A, Foley J, Golla J, McRorie E, Ralston SH. Methotrexate continuation increases fracture risk in patients who sustained lower limb insufficiency fractures. Ann Rheum Dis 2025.
[2] Ragab AH, Frech RS, Vietti TJ. Osteoporotic fractures secondary to methotrexate therapy of acute leukemia in remission. Cancer 1970;25:580–5. doi: 10.1002/1097-0142(197003)25:3<580::aid-cncr2820250313>3.0.co;2-m.
[3] Preston SJ, Diamond T, Scott A, Laurent MR. Methotrexate osteopathy in rheumatic disease. Ann Rheum Dis 1993;52:582–5. doi: 10.1136/ard.52.8.582.
[4] Rolvien T, Jandl NM, Stürznickel J, Beil FT, Kötter I, Oheim R, et al. Clinical and radiological characterization of patients with immobilizing and progressive stress fractures in methotrexate osteopathy. Calcif Tissue Int 2021;108:219–30. doi: 10.1007/s00223-020-00765-5.
[5] Ruffer N, Krusche M, Beil FT, Amling M, Kötter I, Rolvien T. Clinical features of methotrexate osteopathy in rheumatic musculoskeletal disease: a systematic review. Semin Arthritis Rheum 2022;52:151952. doi: 10.1016/j.semarthrit.2022.151952.
[6] Hoenig T, Ackerman KE, Beck BR, Bouxsein ML, Burr DB, Hollander K, et al. Bone stress injuries. Nat Rev Dis Primers 2022;8:26. doi: 10.1038/s41572-022-00352-y.
[7] Ruffer N, Stürznickel J, Rolvien T. Bilateral stress fractures of the lower extremity in methotrexate osteopathy. Dtsch Ärztebl Int 2022;119:231. doi: 10.3238/arztebl.m2022.0019.
[8] Rolvien T, Creutzfeldt AM, Lohse AW, Amling M. Stress fractures in systemic lupus erythematosus after long-term MTX use successfully treated by MTX discontinuation and individualized bone-specific therapy. Lupus 2019;28:790–3. doi: 10.1177/0961203319841434.
[9] Leder BZ, Tsai JN, Uihlein AV, Wallace PM, Lee H, Neer RM, Burnett-Bowie SA. Denosumab and teriparatide transitions in postmenopausal osteoporosis (the DATA-Switch study): extension of a randomised controlled trial. Lancet 2015;386:1147–55. doi: 10.1016/S0140-6736(15)61120-5.
[10] Saag KG, Petersen J, Brandi ML, Karaplis AC, Lorentzon M, Thomas T, et al. Romosozumab or alendronate for fracture prevention in women with osteoporosis. N Engl J Med 2017;377:1417–27. doi: 10.1056/NEJMoa1708322.
[11] Hauser B, Raterman H, Ralston SH, Lems WF. The effect of anti-rheumatic drugs on the skeleton. Calcif Tissue Int 2022;111:445–56. doi: 10.1007/s00223-022-01001-y.
[12] May KP, Mercill D, McDermott MT, West SG. The effect of methotrexate on mouse bone cells in culture. Arthritis Rheum 1996;39:489–94. doi: 10.1002/art.1780390317.
[13] King TJ, Georgiou KR, Cool JC, Scherer MA, Ang ES, Foster BK, Xu J, Xian CJ. Methotrexate chemotherapy promotes osteoclast formation in the long bone of rats via increased pro-inflammatory cytokines and enhanced NF-κB activation. Am J Pathol 2012;181:121–9. doi: 10.1016/j.ajpath.2012.03.037.
[14] Paccou J, Yavropoulou MP, Naciu AM, Chandran M, Messina OD, Rolvien T, et al. Prevention and treatment of glucocorticoid-induced osteoporosis in adults: recommendations from the European Calcified Tissue Society. Eur J Endocrinol 2024;191:G1–17. doi: 10.1093/ejendo/lvae146.
[15] Roca-Ayats N, Balcells S, Garcia-Giralt N, Falcó-Mascaro M, Martinez-Gil N, Abril JF, et al. GGPS1 mutation and atypical femoral fractures with bisphosphonates. N Engl J Med 2017;376:1794–5. doi: 10.1056/NEJMc1612804.

Ann Rheum Dis 84 (2025) 521–528

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Viewpoint

How to treat undifferentiated arthritis today or tomorrow? A consideration of treatment recommendations in light of current evidence

Annette H.M. van der Helm-van Mil[1,*], Daniel Aletaha[2]

[1] *Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands*
[2] *Department of Internal Medicine III, Division of Rheumatology, Medical University of Vienna, Vienna, Austria*

ARTICLE INFO

Article history:
Received 23 September 2024
Received in revised form 3 December 2024
Accepted 7 January 2025

ABSTRACT

Patients with undifferentiated arthritis (UA) have clinically apparent inflammatory arthritis but no evident diagnosis or classification. Considering UA as a definition 'per exclusionem' implies that the population designated by this term is affected by changes in the way other diseases, eg, rheumatoid arthritis (RA), are classified or diagnosed. Current treatment recommendations for UA are largely similar to recommendations for RA. The recommendations are based on the idea that UA is an early stage of RA and on literature generated in the 2000s before the development of the 2010 classification criteria for RA. However, conventional UA (so-called '1987-UA') is presumably different than contemporary UA ('2010-UA'). Strikingly, there are no randomised placebo-controlled trials done on '2010-UA,' and this poses questions on whether the recommendations for UA are still valid. In this absence, we assume that treatment recommendations from '1987-UA' can be extrapolated to '2010-UA' if (1) essential patient characteristics are the same, (2) long-term outcomes are similar, (3) prognostic factors are largely the same, and (4) there are indications from research other than placebo-controlled randomized clinical trials (RCTs) that disease modifying antirheumatic drug (DMARD) treatment in 2010-UA is effective. We evaluate these requirements one by one based on the literature on 2010-UA. This reveals that 2010-UA is milder in initial presentation and disease outcomes than 1987-UA. Today's UA population is >95% anticitrullinated protein antibody-negative, presents with mono- or oligoarthritis, frequently achieves spontaneous remission, and rarely progresses to RA. We suggest that 2010-UA is a distinct patient group within the early arthritis spectrum, requiring additional research, after which recommendations may need to be updated.

INTRODUCTION

Undifferentiated arthritis (UA) is a concept we think we know. Patients present with clinically apparent inflammatory arthritis (joint swelling at physical examination) and have no clear diagnosis or classification. It is, therefore, also called unclassified arthritis. We recognise that this population is, to some extent, heterogeneous by the nature of the potential underlying diseases it may evolve to, the uncertainty about self-remittance, and the unclear amount of investigative effort needed to robustly claim 'unclassifiability.' Though treatment recommendations are clear and included in the 2016 European Alliance of Associations for Rheumatology (EULAR) recommendations for early arthritis [1]. These state that (1) if a definite diagnosis cannot be reached and the patient has early UA, risk factors for persistent and/or erosive disease, including the number of swollen joints, acute phase reactants, rheumatoid factor (RF), anticitrullinated protein antibodies (ACPAs), and imaging

*Correspondence to Prof A H M van der Helm-van Mil, Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands.
E-mail address: AvdHelm@lumc.nl (A.H.M. van der Helm-van Mil).
Handling editor Prof Kimme Hyrich.

https://doi.org/10.1016/j.ard.2025.01.036

findings, should be considered in management decisions, (2) patients at risk of persistent arthritis should be started on disease modifying antirheumatic drugs (DMARDs) as early as possible (ideally within 3 months), even if they do not fulfil classification criteria for inflammatory rheumatologic disease, (3) among the DMARDs, methotrexate is the anchor drug and, unless contraindicated, should be part of the first treatment strategy in patients at risk of persistent disease, and (4) the main treatment goal is to achieve clinical remission, assessed with regular monitoring of disease activity. Thus, the EULAR recommendations endorse treatment for UA that is very similar to treatment for rheumatoid arthritis (RA).

It is known that recommendations are intended as support and that deviations from these recommendations can be made for individual patients. Because recommendations are generally based on scientific data, the question of how to manage UA, therefore, does not appear to be a recent or pressing issue. The question is, is that true? Are current recommendations still sufficient for patients with UA today or tomorrow? We will here evaluate the concept of UA, its evolvement over time, and the evidence we have regarding prognostication and intervention.

UA AND ITS HISTORY

The first naming of the term UA occurred in the late 1980s [2]. It was described as a group of patients with early arthritis for whom no clear diagnosis could be made. In that respect, the description of the entity is not substantially different from the current description. It was and still is clearly a diagnosis by exclusion. Over the past 30 years, classification criteria for several inflammatory arthritides have changed, of which changes in the classification criteria for RA may have the most numerical impact on UA. Although classification criteria are not intended as a diagnostic tool, these changes do influence the general concept of a disease.

In the first description of UA in the late 1980s, RA was excluded by referring to the 1958 ARA classification criteria. The number of studies on UA, which reflects interest in UA, began to increase from 2003 to 2005, with a peak in the number of publications in 2011. These figures suggest that the majority of studies on UA were conducted when RA was defined according to the 1987 classification criteria. A suggestion that is confirmed when reading the studies in more detail. The vast majority of studies on prognostic markers studied a UA population that had no clear diagnosis, including RA, according to the 1987 criteria. A widely used and extensively validated prediction model that estimated the risk of RA was also derived for the '1987-UA-population' [3−5]. The placebo-controlled randomised clinical trials done in UA are summarised elsewhere [6]. There are 5 trials on UA; all were designed and executed before 2010. Consequently, all trials were done in a 1987-UA population (except for the PROMPT trial, which used an even older definition and studied 1958-UA, at that time called 'possible RA'). The primary endpoint in all of these trials was the development of RA (according to clinical diagnosis without or with fulfilment of the 1987 criteria for RA).

As is known, the classification criteria for RA were changed again in 2010, with the aim of being able to recognise or classify RA earlier. Some of the patients that were formerly classified as UA are now characterised as RA; this mostly concerned ACPA-positive patients [7]. This has obviously changed the remaining group of UA patients (that we will refer to as '2010-UA' or 'contemporary UA').

DIAGNOSING UA

In clinical practice, we do not consider classification criteria (eg, for RA) to exclude other diagnoses and thereby identify patients with UA. There is also no recommendation on what diagnostic test should be done or what classifiable diseases should be ruled out before clinical arthritis can be labelled as undifferentiated. In daily practice, UA is therefore diagnosed by clinical expertise. A recent study covering UA patients diagnosed between 1993 and 2019 studied 'expertise-based UA' and showed that this UA population has changed over time [8]. This study illustrates that classification criteria influence what doctors consider a disease entity. More importantly, these results hint at the fact that we must pay close attention to whether the knowledge from the literature applies to today's patient group.

HOW TO MANAGE 2010-UA?

There is a paradox between the scientific data being largely based on the 1987-UA and the population of UA patients that we see in daily practice. Also, the above-mentioned EULAR recommendations for UA are based on studies that evaluated populations with early RA or early arthritis, and none of the studies referred to were done on 2010-UA [1]. Also, no placebo-controlled randomised clinical trials have been conducted on '2010-UA'/'contemporary UA.' The question is, therefore, whether the results of the prognostic studies and clinic trials in the 1987-UA population can be extrapolated to the current 2010-UA population.

In the absence of placebo-controlled randomised clinical trials in 2010-UA, we presume that the data from 1987-UA and recommendations for UA can be extrapolated to 2010-UA if (1) essential patient characteristics are the same, (2) long-term outcomes are similar, (3) prognostic factors are largely the same, and (4) there are indications from research other than placebo-controlled randomized clinical trials (RCTs) that DMARD treatment in 2010-UA is effective. We will consider these requirements one by one.

PATIENT CHARACTERISTICS HAVE CHANGED

Some classic characteristics that describe a rheumatologic patient population are the number of swollen and tender joints, their distribution at physical examination, and autoantibodies (ACPAs and RF) as laboratory results. Approximately 25% to 65% of 1987-UA patients are ACPA-positive [4,9−12]. Due to the composition of the 2010 classification criteria for RA, in which autoantibodies carry a heavy weight, ACPAs have become rare in 2010-UA. Only 4% to 5% of 2010-UA patients are ACPA-positive (Fig 1) [3,4,9−16].

The number of tender and swollen joints has also changed. The mean number of tender joints in 1987-UA ranged between 5 and 7, and the number of swollen joints between 3 and 6 [4,9−12]. In 2010-UA, in contrast, patients presented with, on average, 1 or 2 tender joints [13,14] and also 1 or 2 swollen joints (Fig 1) [8,13−16].

Other features, such as age, gender, and the frequency of increased acute phase reactants, remained largely unchanged. Hence, the majority of current UA patients are ACPA-negative and present with mono- or oligoarthritis (Fig 1).

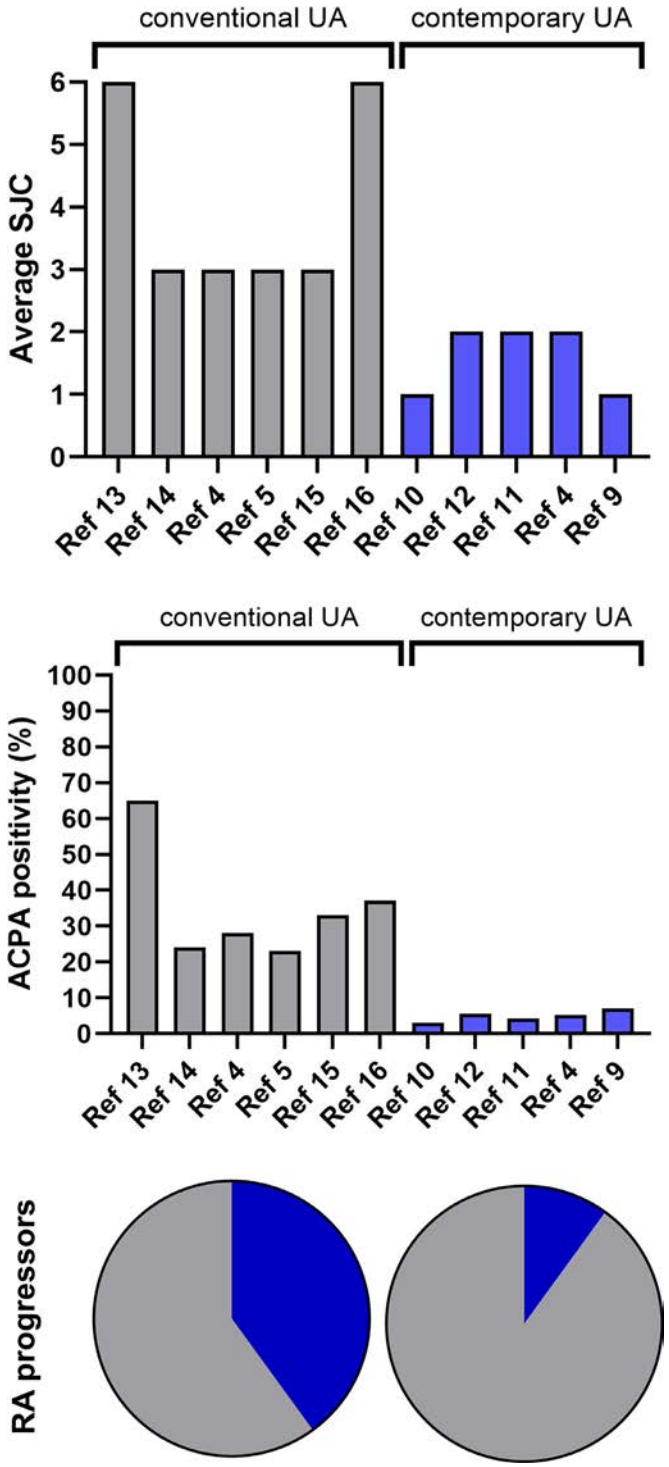


Figure 1. Average number of swollen joints (upper), percentage anticitrullinated protein antibody (ACPA) positivity (middle) , and percentage of patients progressing to RA (lower) in patients presenting with conventional undifferentiated arthritis (UA) ('1987-UA') and contemporary UA ('2010-UA'). The figures on the swollen joint count (SJC) and ACPA positivity are based on data from [3,4,9–16]. The percentage of UA patients that developed RA is indicated in blue and based on data from [4,14].

LONG-TERM OUTCOMES HAVE CHANGED

It is not surprising that as current UA comprises a different patient population, long-term outcomes differ as well. Compared with a previous rate of conversion to RA of 30% to 45% in 1987-UA [3,4], ≤10% of 2010-UA develops RA (according to the 2010 criteria) (Fig 1). RA development generally occurs within the first year [14]. Whether 1987-UA and 2010-UA differ in the course of disease activity score (DAS) and functional disabilities over time remains to be studied. Interestingly, a recent study revealed that about 60% of 2010-UA patients achieved sustained remission without DMARDs. This was defined as the sustained absence of swollen joints and included not only DMARD-free sustained remission but also spontaneous sustained remission (without ever DMARD treatment); this might be considered as resembling a cure [8]. Furthermore, in contrast to RA and 1987-UA, patients with 2010-UA have no excess mortality [13]. So, although some long-term outcomes remain to be studied in 2010-UA, available data suggest that the course of 2010-UA is more favourable than for 1987-UA, which in turn is also confirmatory that the 2010 RA criteria are capturing more patients with a poor prognostic phenotype: the 'RA phenotype.'

PROGNOSTIC FACTORS HAVE PARTLY CHANGED

Classic prognostic factors

Prognostic factors are generally studied with RA development as an outcome. The EULAR recommendation mentions swollen joints, acute phase reactants, RF, and ACPAs as key prognostic factors [1]. Radiographic erosions have also been shown to be predictive in 1987-UA [17]. Several of these classic risk factors are less informative in 2010-UA. ACPAs are still predictive, but considering their frequency of <5%, it is less clinically relevant [14,15,18]. Raised acute phase reactants (C-reactive protein (CRP) and Erythrocyte sedimentation rate (ESR)) are no longer predictive of RA development [14,15]. Polyarthritis remains a predictor for RA development but has also become infrequently present [14,15]. In addition, its value is slightly dependent on the outcome. When considering disease persistence as a long-term outcome, polyarthritis was no longer an independent risk factor [13]. Radiographic erosions are also not predictive in 2010-UA [17,18]. Thus, several classic risk factors are no longer relevant in 2010-UA. The presence of ACPAs remains the most consistent

Table 1
Prognostic factors in conventional undifferentiated arthritis (1987-UA) and contemporary undifferentiated arthritis (2010-UA)

Prognostic factor	Outcome: development of RA		Outcome: persistent arthritis
	1987-UA	2010-UA	2010-UA
Polyarthritis	+	+	-
ACPA positivity	+	+	+
Increased CRP	+	-	+
Increased ESR	+	-	-
Bone erosion at x-ray	+	-	-
HLA-SE	+	-	NA
Tenosynovitis at MRI	+	+	NA
Anti-Carp antibodies[a]	+	-	NA

ACPA, anticitrullinated protein antibody; +, association with outcome/increased risk; Anti-Carp, anticarbamylated protein antibodies; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; HLA-SE, HLA shared epitope alleles; MRI, magnetic resonance imaging; NA, not assessed; 1987-UA, conventional UA; -, no association/no risk factor; RA, rheumatoid arthritis; 2010-UA, contemporary UA; UA, undifferentiated arthritis.
This table summarises the risk factors described and referenced in the text. The data for HLA-SE alleles in the 2010-UA are in the Supplementary material.
[a] Predictive value is independent of ACPAs and rheumatoid factor.

predictor but can be considered of limited value because of its absence in ~95% of UA patients. An overview of 'classic' and 'novel' risk factors and their predictive ability for RA development in 1987-UA and 2010-UA is provided in Table 1.

Antimodified protein antibodies

Now that classic risk factors are less accurate or helpful, new risk factors were sought, such as newer antimodified protein antibodies (AMPAs). Anticarbamylated protein (Anti-Carp) antibodies were predictive of developing RA in 1987-UA, and that effect was independent of ACPAs and RF. In 2010-UA, however, anti-Carp antibodies were no longer predictive for the development of RA (Table 1) [19]. Anti-Carp was also rare in ACPA-negative, RF-negative UA: 2.5% of seronegative 1987-UA patients were anti-Carp positive, and only 0.01% etc of seronegative 2010-UA [19]. Antibodies against acetylated vimentin have been described [20] but, to the best of our knowledge, have not been tested in 2010-UA. Hence, multiple studies, including various AMPAs, still remain to be done in 2010-UA.

Imaging

Imaging can be more sensitive in detecting joint inflammation than physical examination of joints. Ultrasound (US) and Magnetic Resonance Imaging (MRI) were evaluated in a few studies in 2010-UA. US-detected inflammation was found predictive in 2010-UA [21,22], but regular clinical and serologic risk factors were not studied, and the added value of US remains unidentified.

A Japanese study in 2010-UA provided the first evidence that tenosynovitis detected by MRI is predictive for developing RA [23]. A large study in >400 consecutive 2010-UA patients demonstrated that MRI-detected tenosynovitis was associated with RA progression, independent of regular clinical and serological predictors (Table 1) [15]. A flowchart revealing the value of hand and foot MRI in subgroups of UA patients based on ACPA status and joint involvement revealed that MRI is most valuable in ACPA-negative UA patients with oligoarthritis; the absence of MRI-detected tenosynovitis contributed to excluding future RA development [15]. MRI was less helpful in ACPA-negative monoarthritis as the prior risk for RA was already low [15]. Follow-up research showed that an MRI of one hand is sufficient, that additional scans of the feet are not of added benefit [24], and that MRI-detected erosions are not predictive [25]. Thus, performing a hand MRI in ACPA-negative UA with oligoarthritis could aid in preventing overtreatment by identifying a group of patients that will not develop RA.

Genetics

Human leucocyte antigen (HLA)-shared epitope alleles are the strongest genetic risk factor for RA and ACPA-positive RA, in particular. While predictive in 1987-UA [26,27], the HLA-shared epitope alleles do not confer risk for RA in 2010-UA (Table 1), possibly because some of the genetic risk is already captured by the ACPA positivity.

Risk stratification models

While we have so far looked at the value of individual risk factors, a combination of factors will ultimately be required for accurate risk stratification. The so-called 'Leiden prediction model' was designed for 1987-UA and extensively validated internationally [4,5]. In addition to autoantibodies and the number of involved joints, this model contains information about morning stiffness, distribution of involved joints, symmetry, age, and gender. The idea was that with the advent of the 2010 classification criteria, which also aimed at earlier recognition of RA, this model has become redundant. Interestingly, a number of research groups evaluated the accuracy of this model in 2010-UA and found a discriminative ability that was not much inferior to what is known from the model in 1987-UA (Area under the receiver operator curve (AUC), 0.85; 0.83 in 2010-UA vs 0.87 originally in 1987-UA) [28,29]. Likewise, the evaluation of consecutive 2010-UA patients from the Leiden early arthritis cohorts revealed an AUC of 0.81 (Supplementary File S2). These data suggest that the Leiden prediction model, which is easily filled at MDCalc [30], still has some value in 2010-UA.

What is yet unexplored is whether there are variables in this model that no longer contribute to the prediction in 2010-UA and, therefore, may be omitted. Vice versa, other factors may need to be included; for instance, imaging-detected tenosynovitis may be useful to explore [15].

Finally, a crucial question is whether the development of RA is the pivotal outcome when developing risk stratification methods for UA. Alternatives include persistent arthritis or, vice versa, spontaneous remission. The latter would identify the people who should not be treated. Functional disability and loss of productivity are also of great importance. These outcomes are virtually nonexistent in prognostic research in 2010-UA.

EVIDENCE FOR THE EFFICACY OF DMARD TREATMENT FROM RESEARCH OTHER THAN PLACEBO-CONTROLLED RCTS IS LACKING

Since placebo-controlled clinical trials in 2010-UA are absent, a recent study explored the efficacy of DMARD treatment in 2010-UA using 25 years of observational data in which the inclusion period (periods between 1993 and 2019) was used as an instrumental variable for changes in treatment strategies [8]. It was hypothesised that if the use of DMARD treatment, which had increased over time, had a disease-modifying effect on UA, this resulted in lower disease activity, improved physical functioning, increased frequency of prolonged DMARD-free status, and less progression to RA. Despite the fact that DAS slightly improved in patients diagnosed from 2011 onwards (approximately −0.20 DAS units), which is related to more frequent DMARD treatment functional disability, the prevalence of prolonged DMARD-free status and progression to RA did not concomitantly improve compared with patients diagnosed from 1993 to 1997 [8]. Notably, the observed improvement in DAS28-CRP from 2011 onwards does not exceed the minimal clinically important difference of 1.0. This increased use of therapeutics without notable improvement in long-term outcomes also raises questions about benefits/risks in the treatment of UA and, as such, questions how current UA should be treated.

A possibility to gain some insight into the efficacy of DMARD treatment in 2010-UA in the short term may be to reanalyse the existing UA trials and perform posthoc analyses in which the 2010-UA patients are identified retrospectively and the treatment efficacy explored.

2010-UA IS DIFFERENT FROM PREARTHRITIS OR PRE-RA

With the recent focus on the prevention of RA, the interest in 'prearthitis' and 'pre-RA' has increased. How does 2010-UA fit

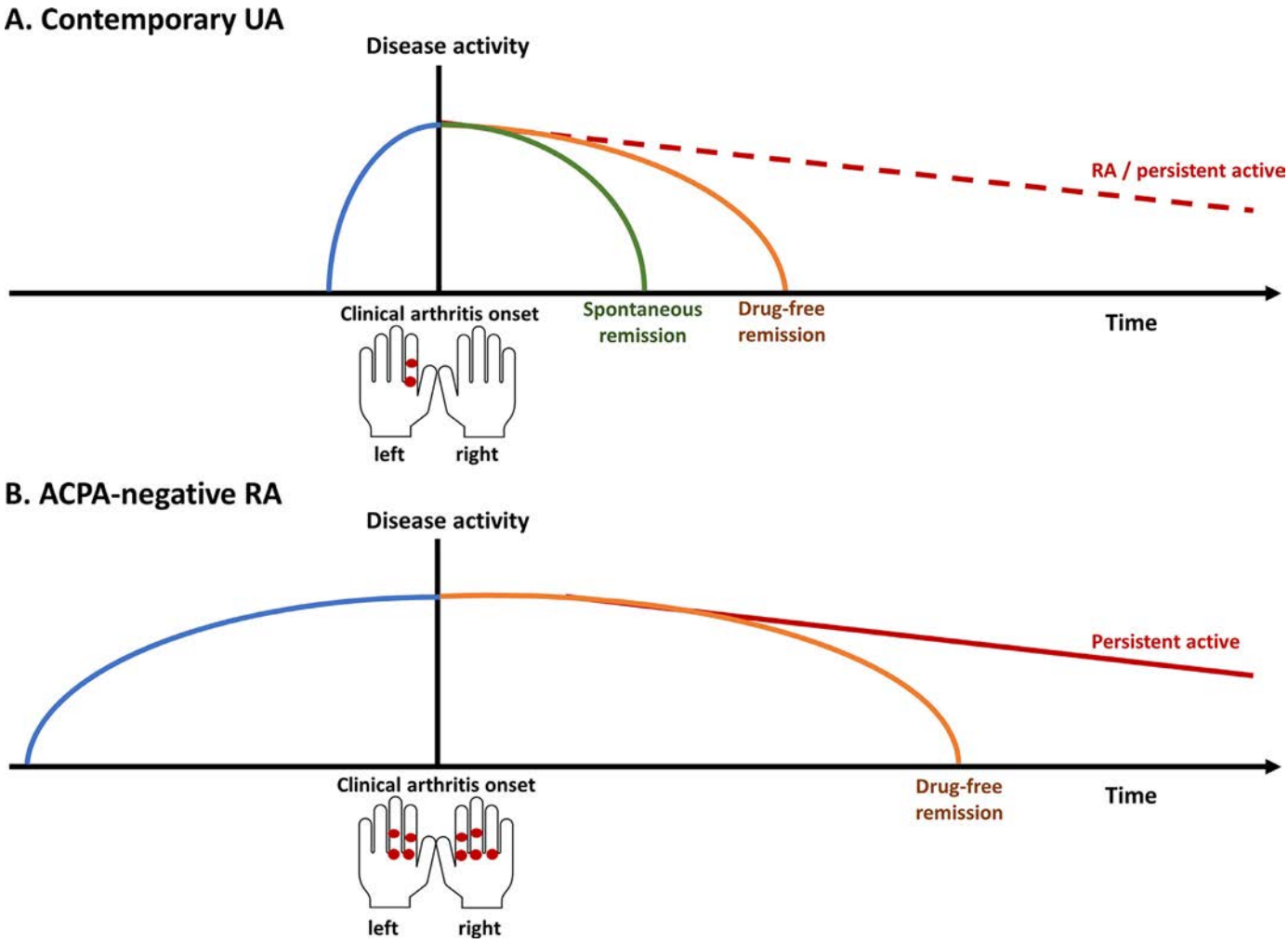


Figure 2. Conceptual differences between autoantibody-negative rheumatoid arthritis (RA) and contemporary undifferentiated arthritis (UA). Autoantibody-negative RA typically has a symptomatic prearthritic phase, presents with polyarthritis of small joints, and has a protracted disease course that may result in sustained Disease Modifying Antirheumatic Drug-free remission (DMARD-free remission) in a subset of patients. Autoantibody-negative RA also has a longer symptom duration at diagnosis than contemporary UA (median, 12-14 weeks in anticitrullinated protein antibody (ACPA)-negative RA [32,33] vs 8 weeks in contemporary UA [16]). Contemporary UA is most frequently autoantibody-negative, presents with fewer swollen joints, and regularly results in spontaneous remission or DMARD-free remission. About 10% progresses to RA. Whether UA has a symptomatic prearthritis stage is unknown. Refining the clinical and pathophysiological differences between contemporary UA and autoantibody-negative RA is part of the research agenda.

into these concepts? The crucial difference between prearthritis and UA is that clinically apparent inflammatory arthritis is, by definition, absent in prearthritis and mandatory for UA. UA has long been seen as a 'pre-RA stage' that occurs just before RA development. However, given the low frequency of 2010-UA that progresses to RA (<10%), this does not apply to the majority of UA patients. This is also supported by findings from arthralgia cohorts. Individuals at risk for RA are often identified by ACPA positivity; once clinical arthritis has developed, RA is diagnosed. Also, people with clinically suspect arthralgia generally progress directly to RA and are rarely first diagnosed with UA at the time of developing clinically apparent inflammatory arthritis [31].

2010-UA AS AN ENTITY IN THE EARLY ARTHRITIS SPECTRUM

The concepts of early arthritis and UA differ by nature; where UA is a specific category of arthritis, and the type of arthritis is unclear, early arthritis obviously refers to the time frame of the disease's onset, regardless of its specific type of arthritis.

Due to the lack of specified tools or criteria to diagnose UA, it is a less defined disease entity than other classified arthritides. Because of the often favourable disease course, an overlap with reactive arthritis could be assumed. However, the absence of obvious infectious episodes preceding autoimmunity makes this speculative. Because of ACPA negativity, an overlap with ACPA-negative RA can be considered. Some clinical characteristics at diagnosis differ between ACPA-negative 2010-UA and ACPA-negative RA at group levels: UA has a shorter symptom duration and a lower number of inflamed joints. Whether UA has a 'pre-UA period' with gradual onset of autoimmune or autoinflammatory phenomena, similar to the risk stage of arthralgia that can precede ACPA-negative RA, remains to be elucidated. Further, the disease course differs; while 2010-UA frequently achieves sustained spontaneous remission or DMARD-free remission (resembling a monocyclic course of joint inflammation), ACPA-negative RA typically has a persistent course that requires DMARD treatment, whereby part of the population can achieve sustained DMARD-free remission over time. Conceptual differences between ACPA-negative RA and 2010-UA are depicted in Figure 2 [16,32,33]. Ideally, in the future, the distinction between UA and RA will be based on differences in underlying

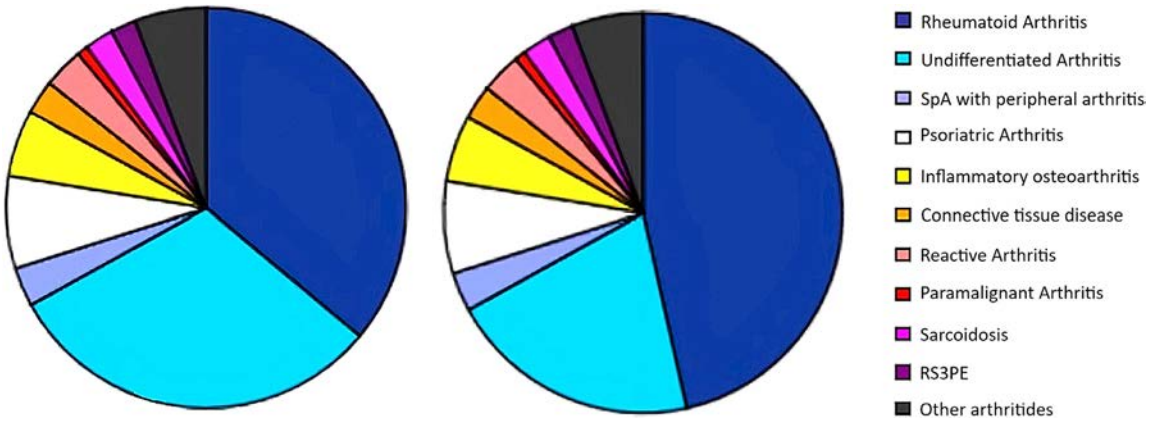


Figure 3. Distribution of diagnosis of consecutive patients with early arthritis in whom rheumatoid arthritis (RA) is defined according to the 1987 criteria (left) and the 2010 criteria (right), which has consequences for the occurrence of undifferentiated arthritis. The data are based on 4369 early arthritis patients who were consecutively included in the Leiden Early Arthritis cohort. Incident early arthritides aged ≥18 years were included, and susception of traumatic arthritis or crystal arthropathy at first presentation were exclusion criteria. Data on diagnoses were collected during the first year of follow-up. RA was defined according to the 1987 criteria (left) or 2010 criteria (right). SpA: Spondyloarthritis; RS3PE: Remitting seronegative, symmetric synovitis with pitting edema.

Table 2
Research agenda

With respect to risk stratification
- Validate the predictive value of MRI-detected tenosynovitis in 2010-UA, as well as an added value to regular clinical and serological predictors.
- Determine the added predictive value of several AMPAs in 2010-UA, in addition to regular clinical and serological predictors.
- Determine the predictive value of genetic variants in 2010-UA, also next to regular clinical and serological predictors.
- Determine the predictive value of ultrasound-detected joint inflammation in 2010-UA, in addition to regular clinical and serological predictors.
- Validation of predictors in independent cohorts from different countries.
- Develop and validate risk stratification algorithms specifically for 2010-UA by evaluating known clinical, serological, imaging, and genetic risk factors.
- Determine whether the outcome of stratification algorithms would include the persistence of arthritis, spontaneous resolution, functional disability, or work loss (rather than RA development).

With respect to the disease course
- Investigate whether 2010-UA is different from 1987-UA in the course of DAS over time and the course of functional limitations.
- Study how 2010-UA differs from 1987-UA in the ability to achieve a DMARD-free status.

With respect to understanding pathobiology

Conduct translational research at the level of systemic markers and synovial tissue:
- To understand the processes underlying 2010-UA
- To unravel heterogeneity within 2010-UA
- To discover the differences between 2010-UA and autoantibody-negative RA

With respect to differentiation/classification
- Assess whether current methods to differentiate 2010-UA from autoantibody-negative RA suffice or need to be redefined.

With respect to treatment
- Perform posthoc analyses on published trials in 1987-UA, select the 2010-UA patients, and evaluate treatment efficacy.
- Perform randomised clinical trials on 2010-UA (not only with RA development as an outcome but also disease persistence and functional disability over time).
- Determine whether methotrexate is effective as first-line DMARD in 2010-UA.
- Determine whether treat-to-target is effective in 2010-UA.
- Summarise the results of future research and update treatment recommendations for UA.

AMPA, antimodified protein antibody; DAS, disease activity score; DMARD, disease modifying antirheumatic drug; MRI, magnetic resonance imaging; 1987-UA, conventional UA; RA, rheumatoid arthritis; 2010-UA, contemporary UA; UA, undifferentiated arthritis.

biological mechanisms, which are, however, not yet fully understood.

Therefore, despite its less well-defined definition and unknown aetiology, UA, in its contemporary form, has become an entity within the spectrum of early arthritis diagnoses. It presents as an ACPA-negative mono- or oligoarthritis and may have a less persistent disease course than several classified arthritides. Figure 3 shows the lower prevalence of 2010-UA compared with 1987-UA in the spectrum of early arthritides.

UNRAVELLING HETEROGENEITY AMONG 2010-UA

Although 2010-UA is a separate entity, heterogeneity may still exist. Unravelling heterogeneity within UA at the pathophysiological level is still impossible. Some studies have examined inflammatory cells in the systemic circulation and synovial tissue of 2010-UA and reported on differences in macrophage density or expression of monocyte-related markers in relation to the clinical outcome [34–36]. Yet, little translational research has been conducted on 2010-UA.

Alternatively, the heterogeneity at the level of clinical variables was unravelled. A unique set of 2010-UA patients, with longitudinal data collected in the era before DMARD treatment became common, were studied for the natural disease course [37]. Clusters were searched based on clinical features at initial presentation, resulting in 5 clusters differentiating patients mostly on joint involvement (3 clusters with monoarthritis, 1 with oligoarthritis, and 1 with polyarthritis). Thereafter, the clusters were evaluated by studying the disease course. The first monoarthritis cluster (18% of UA patients) had clinical arthritis of a large joint, often an acute/subacute onset of symptoms, and were more

often men. Another cluster (9%) concerned monoarthritis of the wrist; these patients were frequently obese, had rapid onset, and had no morning stiffness. The third concerned monoarthritis of a small hand or foot joint (10% of the UA population); this cluster had few characteristic clinical elements, but patients were relatively young (<50 years) and often female. The other 2 clusters were characterised by initial presentation with oligoarthritis (43% of the UA population) or polyarthritis (20% of the UA population) [37]. Patients with unfavourable outcomes (either progression to RA or having persistent disease) were hardly included in the 3 monoarthritis clusters [37]. It could be assumed that DMARD treatment would add little to these monoarthritic clusters. Follow-up research is needed to evaluate this.

CONCLUSION ON THE MANAGEMENT OF 2010-UA

All in all, the 2010-UA population is milder, both in initial presentation and in the severity of the disease course, than in 1987-UA. Today's UA population is >95% ACPA-negative, presents with mono- or oligoarthritis, frequently achieves spontaneous remission, and rarely (about 10%) progresses to RA. 2010-UA, therefore, appears to be its own patient group or entity within the spectrum of early arthritis. Research is needed on risk stratification and appropriate treatment in this patient group (see the research agenda in Table 2). It is risky that the current EULAR guidelines on the treatment of UA appear to be based on the idea that UA is largely similar to RA. Until there is more evidence that can be summarised in renewed EULAR recommendations, the treatment of 2010-UA is a matter of common sense, taking into account the risk of undertreatment and overtreatment, as well as the wishes and possible treatment goals of the patient.

Acknowledgements

Dennis Ton is acknowledged for this help in designing the figures.

Contributors

AHMvdH-vM drafted the manuscript. AHMvdH-vM and DA finalised the manuscript and approved the final version. AHMvdH-vM is the guarantor.

Funding

None.

Competing interests

None.

Patient consent for publication

Not applicable.

Provenance and peer review

Not commissioned; externally peer reviewed.

Ethics approval

Not applicable.

Data availability statement

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.036.

Orcid

Annette H.M. van der Helm-van Mil: http://orcid.org/0000-0001-8572-1437

REFERENCES

[1] Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Álvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2017;76(6):948–59.
[2] Zeidler H, Werdier D, Klauder A, Brinkmann S, Viswat M, Mones ML, et al. Undifferentiated arthritis and spondylarthropathy as a challenge for prospective follow-up. Clin Rheumatol 1987;6(Suppl 2):112–20.
[3] van der Helm-van Mil AH, le Cessie S, van Dongen H, Breedveld FC, Toes RE, Huizinga TW. A prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: how to guide individual treatment decisions. Arthritis Rheum 2007;56(2):433–40.
[4] van der Helm-van Mil AH, Detert J, le Cessie S, Filer A, Bastian H, Burmester GR, et al. Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: moving toward individualized treatment decision-making. Arthritis Rheum 2008;58(8):2241–7.
[5] McNally E, Keogh C, Galvin R, Fahey T. Diagnostic accuracy of a clinical prediction rule (CPR) for identifying patients with recent-onset undifferentiated arthritis who are at a high risk of developing rheumatoid arthritis: a systematic review and meta-analysis. Semin Arthritis Rheum 2014;43(4):498–507.
[6] Burgers LE, Raza K, van der Helm-van Mil AH. Window of opportunity in rheumatoid arthritis - definitions and supporting evidence: from old to new perspectives. RMD Open 2019;5(1):e000870.
[7] van der Helm-van Mil AH, Zink A. What is rheumatoid arthritis? Considering consequences of changed classification criteria. Ann Rheum Dis 2017;76(2):315–7.
[8] Verstappen M, Matthijssen XME, van der Helm-van Mil AHM. Undifferentiated arthritis: a changing population who did not benefit from enhanced disease-modifying anti-rheumatic drug strategies-results from a 25 year longitudinal inception cohort. Rheumatology (Oxford) 2022;61(8):3212–22.
[9] Saleem B, Mackie S, Quinn M, Nizam S, Hensor E, Jarrett S, et al. Does the use of tumour necrosis factor antagonist therapy in poor prognosis, undifferentiated arthritis prevent progression to rheumatoid arthritis? Ann Rheum Dis 2008 Aug;67(8):1178–80.
[10] Fransen J, Visser K, van Dongen H, Huizinga T, van Riel PL, van der Heijde DM. Validity of the disease activity score in undifferentiated arthritis. Arthritis Care Res (Hoboken) 2010;62(10):1392–8.
[11] Tamai M, Kawakami A, Uetani M, Takao S, Arima K, Iwamoto N, et al. A prediction rule for disease outcome in patients with undifferentiated arthritis using magnetic resonance imaging of the wrists and finger joints and serologic autoantibodies. Arthritis Rheum 2009;61(6):772–8.
[12] Kuriya B, Cheng CK, Chen HM, Bykerk VP. Validation of a prediction rule for development of rheumatoid arthritis in patients with early undifferentiated arthritis. Ann Rheum Dis 2009;68(9):1482–5.
[13] den Hollander NK, Verstappen M, Huizinga TW, van der Helm-van Mil A. Management of contemporary early undifferentiated arthritis: data on EULAR's recommendation on the risk of persistent disease. Ann Rheum Dis 2022;81(5):740–1.
[14] Brinkmann GH, Norli ES, Kvien TK, Haugen AJ, Grøvle L, Nygaard H, et al. Disease characteristics and rheumatoid arthritis development in patients with early undifferentiated arthritis: a 2-year followup study. J Rheumatol 2017;44(2):154–61.
[15] den Hollander NK, Verstappen M, Sidhu N, van Mulligen E, Reijnierse M, van der Helm-van Mil AHM. Hand and foot MRI in contemporary

undifferentiated arthritis: in which patients is MRI valuable to detect rheumatoid arthritis early? A large prospective study. Rheumatology (Oxford) 2022;61(10):3963–73.
[16] Verstappen M, Huizinga TW, van der Helm-van Mil AH. No excess mortality in contemporary undifferentiated arthritis, in contrast to rheumatoid arthritis: a study with a follow-up of at least 10 years. Ann Rheum Dis 2022;81(8):1197–9.
[17] Thabet MM, Huizinga TW, van der Heijde DM, van der Helm-van Mil AH. The prognostic value of baseline erosions in undifferentiated arthritis. Arthritis Res Ther 2009;11(5):R155.
[18] Mueller RB, Kaegi T, Haile SR, Schulze-Koops H, Schiff M, von Kempis J. Clinical and radiographic course of early undifferentiated arthritis under treatment is not dependent on the number of joints with erosions at diagnosis: results from the Swiss prospective observational cohort. RMD Open 2018;4(1):e000673.
[19] Boeters DM, Trouw LA, van der Helm-van Mil AHM, van Steenbergen HW. Does information on novel identified autoantibodies contribute to predicting the progression from undifferentiated arthritis to rheumatoid arthritis: a study on anti-CarP antibodies as an example. Arthritis Res Ther 2018;20(1):94.
[20] Juarez M, Bang H, Hammar F, Reimer U, Dyke B, Sahbudin I, et al. Identification of novel antiacetylated vimentin antibodies in patients with early inflammatory arthritis. Ann Rheum Dis 2016;75(6):1099–107.
[21] Gutierrez M, Bertolazzi C, Castillo E, Reyes-Long S, Clavijo-Cornejo D, Santos-Moreno P. Ultrasound as a useful tool in the diagnosis of rheumatoid arthritis in patients with undifferentiated arthritis. J Clin Rheumatol 2019;25(5):203–8.
[22] Horton SC, Tan AL, Wakefield RJ, Freeston JE, Buch MH, Emery P. Ultrasound-detectable grey scale synovitis predicts future fulfilment of the 2010 ACR/EULAR RA classification criteria in patients with new-onset undifferentiated arthritis. RMD Open 2017;3(1):e000394.
[23] Tamai M, Grundeken V, Arima K, Brinck RT, Mil AHMVH, Ohki N, et al. Predictive value of magnetic resonance imaging-detected tenosynovitis of the metacarpophalangeal and wrist joints for the development of rheumatoid arthritis among patients with undifferentiated arthritis. Intern Med 2023;62(16):2329–34.
[24] Dakkak YJ, Boeters DM, Boer AC, Reijnierse M, van der Helm-van Mil AHM. What is the additional value of MRI of the foot to the hand in undifferentiated arthritis to predict rheumatoid arthritis development? Arthritis Res Ther 2019;21(1):56.
[25] Boeters DM, Nieuwenhuis WP, van Steenbergen HW, Reijnierse M, Landewé RBM, van der Helm-van Mil AHM. Are MRI-detected erosions specific for RA? A large explorative cross-sectional study. Ann Rheum Dis 2018;77(6):861–8.
[26] van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Huizinga TW, Toes RE, de Vries RR. The HLA-DRB1 shared epitope alleles are primarily a risk factor for anti-cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. Arthritis Rheum 2006;54(4):1117–21.
[27] Huizinga TW, Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, Jawaheer D, et al. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum 2005;52(11):3433–8.
[28] Bedran Z, Quiroz C, Rosa J, Catoggio LJ, Soriano ER. Validation of a prediction rule for the diagnosis of rheumatoid arthritis in patients with recent onset undifferentiated arthritis. Int J Rheumatol 2013;2013:548502.
[29] Ghosh K, Chatterjee A, Ghosh S, Chakraborty S, Chattopadhyay P, Bhattacharya A, et al. Validation of Leiden Score in Predicting progression of rheumatoid arthritis in undifferentiated arthritis in Indian population. Ann Med Health Sci Res 2016;6(4):205–10.
[30] MDCalc. Leiden clinical prediction rule for undifferentiated arthritis [Internet]. 2024. Available from: https://www.mdcalc.com/calc/10023/leiden-clinical-prediction-rule-undifferentiated-arthritis
[31] Krijbolder DI, Verstappen M, van Dijk BT, Dakkak YJ, Burgers LE, Boer AC, et al. Intervention with methotrexate in patients with arthralgia at risk of rheumatoid arthritis to reduce the development of persistent arthritis and its disease burden (TREAT EARLIER): a randomised, double-blind, placebo-controlled, proof-of-concept trial. Lancet 2022;400(10348):283–94. doi: 10.1016/S0140-6736(22)01193-X.
[32] Boeters DM, Burgers LE, Sasso EH, Huizinga TWJ, van der Helm-van Mil AHM. ACPA-negative RA consists of subgroups: patients with high likelihood of achieving sustained DMARD-free remission can be identified by serological markers at disease presentation. Arthritis Res Ther 2019;21(1):121.
[33] Verstappen M, Matthijssen XME, Connolly SE, Maldonado MA, Huizinga TWJ, van der Helm-van Mil AHM. ACPA-negative and ACPA-positive RA patients achieving disease resolution demonstrate distinct patterns of MRI-detected joint-inflammation. Rheumatology (Oxford) 2022;62(1):124–34.
[34] Kurowska W, Kuca-Warnawin E, Radzikowska A, Jakubaszek M, Maślińska M, Kwiatkowska B, et al. Monocyte-related biomarkers of rheumatoid arthritis development in undifferentiated arthritis patients - a pilot study. Reumatologia 2018;56(1):10–6.
[35] Cuervo A, Celis R, Julià A, Usategui A, Faré R, Ramírez J, et al. Synovial immunohistological biomarkers of the classification of undifferentiated arthritis evolving to rheumatoid or psoriatic arthritis. Front Med (Lausanne) 2021;8:656667.
[36] Fuentelsaz-Romero S, Cuervo A, Estrada-Capetillo L, Celis R, García-Campos R, Ramírez J, et al. GM-CSF expression and macrophage polarization in joints of undifferentiated arthritis patients evolving to rheumatoid arthritis or psoriatic arthritis. Front Immunol 2020;11:613975.
[37] den Hollander NK, Verstappen M, van Dijk BT, van der Helm-van Mil AHM, van Steenbergen HW. Disentangling heterogeneity in contemporary undifferentiated arthritis - a large cohort study using latent class analysis. Semin Arthritis Rheum 2023;63:152251.

Ann Rheum Dis 84 (2025) 529–537

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Recommendation

EULAR points to consider for patient education in physical activity and self-management of pain during transitional care

Javier Courel-Ibáñez[1], Rafael Prieto-Moreno[2,3,*], Erica Briones-Vozmediano[4,5], Patrocinio Ariza-Vega[6,7,8], Saskya Angevare[9], Jordi Anton[10], Ilaria Bini[11,12], Daniel Clemente[13], Matilde Correia[14], Wendy Costello[9], Diederik De Cock[15], Andrea Domján[16], Leticia Leon[17], Andréa Marques[18], Kirsten Minden[19,20,21], Ana Filipa Mourão[22], Aurélie Najm[23], Seza Ozen[24], Georgina Pimentel[25,26], Zainab Saleem[27], Tomas Vetrovsky[28], Nico M. Wulffraat[29], Andrea Zacarias Crovato[10,30], Yeliz Prior[31], Loreto Carmona[32], Fernando Estévez-López[2,3,33]

[1] *Department of Physical Education and Sports, University of Granada, Granada, Spain*
[2] *Department of Education, Faculty of Education Sciences, CIBIS (Centro de Investigación para el Bienestar y la Inclusión Social) Research Centre and SPORT Research Group (CTS-1024), University of Almería, Almería, Spain*
[3] *Biomedical Research Unit, Torrecardenas University Hospital, Almería, Spain*
[4] *Faculty of Nursing and Physiotherapy, Consolidated Group of Studies in Society, Health, Education and Culture (GESEC), University of Lleida, Lleida, Spain*
[5] *Healthcare Research Group (GRECS), Biomedical Research Institute of Lleida (IRB Lleida- Josep Pifarre foundation), Lleida, Spain*
[6] *Instituto de Investigación Biosanitaria ibs.Granada, Granada, Spain*
[7] *Department of Physiotherapy, Faculty of Health Sciences, University of Granada, Granada, Spain*
[8] *Department of Physical Education and Sports, Faculty of Sport Science, and PA-HELP "Physical Activity for HEaLth Promotion" Research Group, University of Granada, Granada, Spain*
[9] *European Network for Children with Arthritis and Autoinflammatory Diseases, Geneva, Switzerland*
[10] *Paediatric Rheumatology Department, Hospital Sant Joan de Deu, Barcelona, Spain*
[11] *Anmar Young, Rome, Italy*
[12] *EULAR Young PARE, Zürich, Switzerland*
[13] *Paediatric Rheumatology Unit, Hospital Infantil Universitario "Niño Jesús", Madrid, Spain*
[14] *Centro Hospitalar e Universitário de Coimbra EPE, Coimbra, Portugal*
[15] *Biostatistics and Medical Informatics Research Group, Department of Public Health, Vrije Universiteit Brussel, Brussel, Belgium*
[16] *Department of Internal Medicine, Division of Rheumatology, University of Debrecen Faculty of Medicine, Debrecen, Hungary*
[17] *Rheumatology, Hospital Clinico San Carlos, Madrid, Spain*
[18] *Nursing School of Coimbra, Health Sciences Research Unit: Nursing (UICISA:E), Coimbra, Portugal*
[19] *Klinik für Pädiatrie mit SP Pneumologie, Immunologie und Intensivmedizin, Charité Universitätsmedizin Berlin, Berlin, Germany*
[20] *Epidemiology Unit, Deutsches Rheuma-Forschungszentrum Berlin, Berlin, Germany*
[21] *German Center for Child and Adolescent Health (DZKJ), Berlin, Germany*

*Correspondence to Dr. Rafael Prieto-Moreno.
E-mail address: rafapriemor@gmail.com (R. Prieto-Moreno).
Social media: @javiercourel (J. Courel-Ibáñez), @Rafaprieto4 (R. Prieto-Moreno), @AntonJordi68 (J. Anton), @LeticiaLeonM (L. Leon), @drsezaozen (S. Ozen), @tomasvet (T. Vetrovsky), @ProfYelizPrior (Y. Prior), @carmona_loreto (L. Carmona), @FerEstevezLopez (F. Estévez-López)

Handling editor Josef S. Smolen

https://doi.org/10.1136/ard-2024-226448

[22] Nova Medical School, Rheumatology Department, Centro Hospitalar de Lisboa Ocidental EPE, Lisboa, Portugal
[23] School of Infection and Immunity, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK
[24] Pediatric Rheumatology, Hacettepe University, Ankara, Turkey
[25] Rheumatology Department, Unidade Local de Saúde de Coimbra, Coimbra, Portugal
[26] Nursing Research, Innovation and Development Centre of Lisbon (CIDNUR), Higher School of Nursing, Lisboa, Portugal
[27] EULAR Young PARE and Anmar Young, Zürich, Switzerland
[28] Faculty of Physical Education and Sport, Charles University, Praha, Czech Republic
[29] Pediatrics, University Medical Center, Utrecht, Netherlands
[30] Rheumatology, Hospital Clinic de Barcelona, Barcelona, Spain
[31] School of Health and Society, Centre for Human Movement and Rehabilitation, University of Salford, Salford, UK
[32] Instituto de Salud Musculoesquelética (INMUSC), Madrid, Spain
[33] Department of Social and Behavioral Sciences, Harvard T H Chan School of Public Health, Boston, Massachusetts, USA

ARTICLE INFO

Article history:
Received 19 July 2024
Accepted 17 October 2024

Keywords:
Arthritis
Juvenile
Child
Pain
Nursing
Psychology

ABSTRACT

Objectives: A EULAR task force was convened to develop points to consider (PtC) for patient education in physical activity and self-management of pain in young people with juvenile-onset rheumatic and musculoskeletal diseases during transitional care.
Methods: A task force of 26 people from 10 European countries followed the EULAR Standardised Operating Procedures to establish overarching principles (OAPs) and PtC based on a literature review and expert consensus. Level of evidence (LoE), grade of recommendation (GoR) and level of agreement (LoA) were determined.
Results: Two OAPs and seven PtC were formulated. The OAPs highlight the importance of personalised transitional care in rheumatology, ideally based on shared decision-making and incorporate interactive education to empower young individuals in managing their physical activity and pain. The PtC emphasise the clinical importance of patient education in these areas to improve readiness to transfer from paediatric to adult care. For two PtC, the GoR was moderate (grade B), based on individual cohort study (LoE 2b). For the remaining five PtC, the GoR was weak (grade D), based on expert opinion (LoE 5). The LoA among the task force was high, ranging from 9.4 to 9.8, except for one PtC that was 8.7.
Conclusion: These EULAR PtC establish guidance on best practices for delivering patient education in physical activity and self-management of pain during transitional care in rheumatology. The adoption of these PtC in clinical settings is recommended to standardise and optimise transitional care across European healthcare systems. Additionally, the task force expects that these PtC will drive future research and potentially shape policies across Europe.

INTRODUCTION

The transition from adolescence to adulthood is a critical period characterised by rapid and extensive biopsychosocial changes. This transition is particularly challenging for individuals growing up with juvenile-onset rheumatic and musculoskeletal diseases (jRMDs) [1], which may limit them from having a healthy and satisfactory life [2–4]. Children, adolescents and young adults (hereinafter referred to as young people) with jRMDs have a particular medication and face a health burden [5], distinct from adults [6] and require specialised care [5,6]. Although remission of jRMDs is possible [7], still 50% of cases persist with active disease into adulthood, requiring a transition from paediatric to adult care to continue their treatment [8,9]. During this transition, 20%–50% of young people with jRMDs discontinue treatment, leading to worsening outcomes [10–12]. Thus, in 2017, EULAR/PReS launched general standards and recommendations for transitional care in rheumatology [13]. These guidelines highlighted the importance of early access to developmentally appropriate transitional care provided by specialised and coordinated multidisciplinary teams. While the 2017 EULAR/PReS standards have inspired meaningful progress in the field [14–16], our task force identified a clinically relevant and timely gap: the need for specific guidelines focused on patient education in physical activity and self-management of pain.

First, young people with jRMDs often are at risk of worse overall health and an increased risk of comorbidity in adulthood [9]. In this context, physical activity has well-known, broad benefits for health, including those related to cardiovascular health [17], inflammation [18], cancer [19], sleep quality [20], mental health [21] and health-related quality of life [22]. Physical activity has also specific benefits for RMDs, with improvements in disease activity [23] and symptoms [24]. Therefore, patient education promoting regular physical activity is important and particularly timely now that 81% of adolescents worldwide do not meet the minimum physical activity levels recommended by the WHO [25]. Importantly, this figure seems to be even worse in young people with jRMDs [26]. Second, despite good pharmacological control, an average of 60% of young people with jRMDs continue to experience pain [27]. Persistent pain may negatively impact school attendance [28,29], academic performance [30], social participation [31] and employment rates [4,32], all of which may affect adulthood. In this context, patient education that improves self-management of pain may help young people with jRMDs navigate key outcomes such as educational and vocational achievement, emotional management, intimate relationships and social life [33]. Thus, patient education in physical activity and self-management of pain is essential to manage their health proactively as they transition from paediatric to adult care, offering potential long-lasting benefits.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- The 2017 EULAR/PReS standards and recommendations provide general guidance for transitional care in rheumatology.
- The potentially long-lasting benefits of patient education in physical activity and self-management of pain during transitional care are widely acknowledged.
- However, there is currently a lack of consensus on how best to deliver patient education in these areas during transitional care.

WHAT THIS STUDY ADDS

- This EULAR task force establishes two overarching principles and seven points to consider (PtC) for patient education in physical activity and self-management of pain during transitional care.
- The task force also proposed a research agenda to increase evidence-based knowledge and improving practice to enhance the quality of transitional care.
- These PtC provide guidance for clinical practice in these areas, aiming at facilitating continuity of care in the transition from paediatric to adult services.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- This task force encourages future well-designed randomised controlled trials to evaluate existing patient education programmes in physical activity and self-management of pain during transitional care and, if needed, optimise them or develop new programmes.
- Implementing these PtC in practice may lead to innovative approaches to promote physical activity and self-management of pain during transitional care.
- This PtC may influence policy by advocating for quality of care in rheumatology during the transition from paediatric to adult services across Europe.

Therefore, a task force was convened to establish the EULAR overarching principles (OAPs) and points to consider (PtC) for patient education in physical activity and self-management of pain in jRMDs during transitional care. This task force aimed to complement, rather than update, the 2017 EULAR/PReS general standards. The target audience comprises adult and paediatric rheumatologists, health professionals in rheumatology (HPRs), young people with jRMDs, their families, patient organisations and policy-makers across Europe.

METHODS

The work of this task force was conducted in accordance with the EULAR updated Standardised Operating Procedures [34]. After approval, from the EULAR Executive Committee, a steering committee was established, including the convenor (FE-L), two co-methodologists (FE-L and LC) and two fellows (JC-I and RP-M). The multidisciplinary task force (n = 26) included paediatric and adult rheumatologists, HPRs, people with RMDs and representatives from the Emerging EULAR Network from 10 European countries (Belgium, Czech Republic, Hungary, Germany, the Netherlands, Portugal, Spain, Switzerland, Turkey and the UK). All members disclosed their conflicts of interest upfront. All task force meetings were conducted online.

The steering committee convened in March 2021 to set out the whole process, establishing the protocol to develop the PtC. In line with the meeting's agenda, the fellows conducted a scoping review to create a preparatory material offering insights into current practices of transitional care in young people with jRMDs. Previous EULAR recommendations focused on adults and related to the scope of this task force were also reviewed (eg, patient education [35], pain management [36], physical activity [37] and self-management of inflammatory arthritis) [38]. During the first task force meeting, this preparatory material was discussed, with the final aim of defining the research question for the systematic literature review (SLR), a matter for a separate publication. Within this meeting, we identified the clinical gaps, which would serve as the basis for adopting the Population Intervention Comparison Outcome approach. This approach supported the development of the search strategy for the SLR, conducted in different databases (ie, Cochrane Library, Medline). The SLR was conducted by fellows, with the supervision of the convenor and co-methodologists, to collect qualitative and quantitative studies conducting or describing a structured transitional care programme for young people with jRMDs. Although the primary aim of the SLR was to inform the PtC for patient education in physical activity and self-management of pain, we intentionally broadened the scope of the SLR. This approach aligned with the research agenda developed during our first meeting, ensuring a more comprehensive summary of the current state of the art of transitional care in rheumatology. By including a wider range of studies, we aimed not only to address immediate priorities but also to provide insights that may optimise existing strategies and support the development of more effective, evidence-based approaches. Additionally, this broader review may facilitate the implementation of these approaches in clinical practice by identifying gaps and areas for improvement. In the second meeting, the task force discussed the SLR results and the first list of OAPs and PtC, drafted by the steering committee based on the preparatory material and the semistructured interviews carried out individually with each task force member, where relevant findings were discussed and additional literature was identified. The first two Delphi rounds were performed during the second meeting, where each member of the task force was asked to indicate their level of agreement (LoA) on each recommendation, which was rated anonymously through a survey on an 11-point Likert scale from 0 to 10 (0: completely disagree, 10: completely agree). Items scored over 7 mean points were considered for the next round. Items below 7 mean points were further discussed to decide whether to change the wording/contents or to dismiss them from the list. Additionally, related items were discussed to be grouped. After the meeting, a second list of OAPs and PtC was formulated and LoA asked to be rated anonymously. In the third Delphi round, each member of the task force gave their final rating on every item. The consensus was reached if ≥75% of the members voted in favour of the recommendations and gave ≥8 mean points. The mean and SD of the LoA, as well as the percentage of agreement ≥8, were presented. Lastly, the task force was consulted to revise and provide feedback on the final document. The steering committee appraised the level of evidence (LoE) and grade of recommendation (GoR), according to the standards of the Oxford Centre for Evidence-Based Medicine [39].

RESULTS

The task force agreed on two OAPs and seven PtC (table 1). Most PtC were based on expert opinion (GoR = D, LoE = 5) due to the limited evidence available in the literature. However, PtC 5 and PtC 6 were based on individual cohort study (GoR = B, LoE = 2 b). Overall, the LoA on the present PtC was high, with scores ranging from 9.4 to 9.8, except for PtC 4 (LoA = 8.7). The

Table 1
Overarching principles (OAPs) and points to consider (PtC) for patient education in physical activity and self-management of pain in juvenile-onset rheumatic and musculoskeletal diseases (jRMDs) during transitional care in rheumatology

				LoA	
OAPs				**Mean (SD)**	**% with score ≥8**
1	Transitional care has clear personalised aims, ideally based on shared decision-making and periodically monitored by both objective and patient-reported outcomes			9.9 (0.8)	100
2	Patient education is a planned interactive learning process designed to support and enable young people with jRMDs to manage their health and chronic condition and optimise their well-being during transitional care			9.5 (0.8)	100
PtC		**LoE**	**GoR**	**Mean (SD)**	**% with score ≥8**
1	Patient education in physical activity and self-management of pain should be prioritised for newly diagnosed patients and for those in transitional care	5	D	9.6 (0.5)	100
2	Patient education in physical activity and self-management of pain should be tailored and needs-based according to the young people's priorities, preferences, capabilities and resources	5	D	9.6 (0.9)	95
3	Patient education during transitional care should consist of a variety of learning formats, including digital health	5	D	9.4 (1.4)	91
4	Patient education during transitional care should include the evaluation of both young people's health literacy and, most importantly, their behavioural changes	5	D	8.7 (0.8)	87
5	All patients and their caregivers should be offered education on the importance of maintaining a healthy lifestyle to better self-manage jRMDs	2b-5	B	9.8 (0.7)	100
6	Physical activity has health benefits for young people with jRMDs and should be promoted during transitional care	2b	B	9.8 (1.0)	95
7	Rheumatologists and health professionals in rheumatology should consider offering a variety of physical activity formats that align with young people's preferences and disease requirements	5	D	9.6 (0.4)	100

GoR, ranging from A ('consistent level 1') to D ('level 5 evidence'). LoA, ranging from 0 ('completely disagree') to 10 ('completely agree'). LoE, ranging from 1('high quality randomised controlled trials') to 5 ('expert opinion'). 2b = individual cohort study.
GoR, grade of recommendation; LoA, level of agreement; LoE, level of evidence.

task force attributed the lower LoA for PtC 4 to challenges in identifying a feasible and practical method for measuring health literacy in daily clinical practice during transitional care in rheumatology.

Overarching principles

The task force emphasises that both OAPs are relevant to all the PtC. These OAPs stress the importance of personalised and interactive patient education processes that empower young people with jRMDs to actively manage their health during the transition from paediatric to adult care [35]. A biopsychosocial approach is emphasised, considering not only medical but also psychological and social factors to ensure holistic support. These OAPs encourage shared decision-making, aligning care with individual needs, preferences and resources (ie, patient-centred care), ensuring regular monitoring of both objective and patient-reported outcomes [13].

Points to consider

PtC 1: patient education in physical activity and self-management of pain should be prioritised for newly diagnosed patients and for those in transitional care

Improving young people's knowledge and health literacy in physical activity and self-management of pain early in transitional care may be meaningful [40,41]. Thus, the task force suggests to consider providing patient education in these areas as the starting point of transitional care [42,43]. In this context, patient education may promote an active lifestyle and help young people cope with the consequences of pain [44]. Equipping young people with jRMDs with early health literacy and self-care skills may empower them to manage their own health independently [45], achieve self-efficacy and parental independence [46], prevent consequences of sedentarism [47], mitigate lifelong pain-related disability [48] and improve both their social and vocational development [49]. Additionally, providing education in both physical activity and self-management of pain may have a synergist effect, yielding larger effects when combined together than when addressed separately [50].

PtC 2: patient education in physical activity and self-management of pain should be tailored and needs-based according to the young people's priorities, preferences, capabilities and resources

The task force suggests tailoring patient education in physical activity and self-management of pain to meet young people's care needs according to their developmental changes in perceptions, interests and maturity during transitional care [49]. Importantly, young people of the same age may have different self-management skills and readiness to transition [51]. Thus, patient education may consider biopsychosocial and environmental needs over age-based approaches [52]. Accordingly, educational needs for physical activity and self-management of pain may be flexible, regularly revised and adjusted to the evolving challenges young people face during transitional care, such as changes in self-advocacy, self-confidence, school demands, body image and sexuality [53,54]. In this context, introducing activity pacing may help young people with jRMDs manage daily activities more sustainably by adapting to the fluctuating nature of their symptoms, including pain. Activity pacing involves breaking activities into realistic goals to prevent overexertion and avoid the boom-and-bust cycle, where overactivity leads to flare-ups and subsequent decreases in physical activity.

PtC 3: patient education during transitional care should consist of a variety of learning formats, including digital health

The task force suggests offering young people with jRMDs a variety of options, allowing them to choose based on their

preferred learning style. The main modes of delivering patient education explored in transitional care in rheumatology, include combinations of face-to-face or remote interventions with individual or group-based approaches [55]. Gamification strategies (eg, avatars, progress indicators and rewards) [56,57], narrative medicine interventions (storytelling) [58] and peer-mentoring [59] may also be considered, given the intrinsic interest shown by young people. Social media-based programmes [60] and video games [61] often show good acceptability and provide engaging alternatives for promoting physical activity in young people with jRMDs. Smartphone-based pain self-management programmes seem to be both feasible and beneficial for young people with jRMDs [44].

The task force advises education providers to be skilled in the delivery formats preferred by young people with jRMDs and be flexible in adapting their approach as the young person develops [35]. Education providers are also encouraged to engage young people in informed shared decision-making, particularly regarding patient education in physical activity and self-management of pain. When needed, the task force highlights the importance of offering training in these areas to education providers, including but not limited to, rheumatologists, HPRs, families and teachers.

PtC 4: patient education during transitional care should include the evaluation of both young people's health literacy and, most importantly, their behavioural changes

Health literacy, defined as the ability to find, understand and use health-related information, is crucial for making informed decisions regarding physical activity and self-management of pain. In young people with jRMDs, insufficient health literacy is associated with worse outcomes including poorer transition readiness [62] and ineffective self-management, often resulting in parents having a more central role during consultations [63,64]. Enhancing health literacy equips young people to make informed decisions, supporting them to have better control of their disease management. In transitional care, it may be important to equip young people with jRMDs with health literacy skills to critically appraise the quality of information available on the internet [65]. In addition to enhance health literacy, patient education may have potential to promote long-lasting behavioural changes that integrate healthy practices, such as engaging in physical activity and self-managing pain, into daily routines [9,45]. Educational programmes tailored to the individual health concerns and life experiences of young people with jRMDs may be beneficial in increasing health literacy and helping them better understand the potential risks related to their condition [66] while also encouraging sustainable behavioural changes [67].

Therefore, the task force emphasises the importance of patient education in physical activity and self-management of pain during transitional care, focusing not only on improving health literacy but also on supporting long-lasting behavioural changes and their integration into daily routines [67–69].

PtC 5: all patients and their caregivers should be offered education on the importance of maintaining a healthy lifestyle to better self-manage jRMDs

Patient education in physical activity and self-management of pain should raise awareness and promote specific behavioural changes in young people with jRMDs during transitional care, helping them and their caregivers understand the importance of a healthy lifestyle for reducing disease activity, preventing comorbidities and achieving other general benefits such as improved well-being [43,70,71]. In this context, a systematic review [72] concluded that there is evidence supporting that adopting a healthy lifestyle may be beneficial for young people with jRMDs. For instance, nutritional interventions such as omega-3 fatty acids and iron supplementation have been shown to strengthen immune status markers, including cytokines [73] and haemoglobin levels [74]. In addition, educational programmes focused on healthy nutrition for both patients and their parents have been found to promote greater calcium intake [75], which may have long-lasting benefits supported by better bone health [76]. Although the existing evidence is limited, the task force also emphasises the importance of offering guidance on avoiding unhealthy habits such as smoking and alcohol consumption [77–79], as well as other important topics like sleep hygiene [3,80], sexual health [81,82] and healthy body composition [83,84], as addressing these behaviours may further enhance long-term health outcomes. To support this, it may be important guiding families in reducing overprotective behaviours, encouraging a gradual transfer of responsibility to young people, fostering independence, better self-management and improved health outcomes [15,16,85].

PtC 6: physical activity has health benefits for young people with jRMDs and should be promoted during transitional care

EULAR recognises promoting physical activity is a core component of RMDs care [13,37,86] with its health benefits extending to young people with jRMDs [47]. In this context, there is evidence supporting that physical activity interventions may have benefits in functional status (eg, dressing, walking) [87], health-related quality of life [2], pain [88], multimorbidity [71] and disability [89]. The task force suggests that the WHO 2020 guidelines on physical activity and sedentary behaviour are also applicable to young people with jRMDs during transitional care [25]. We also highlight that, to promote adherence to physical activity, strategies may address barriers faced by young people with jRMDs (eg, fear of worsening symptoms, parental overprotection and low self-efficacy) while maximising facilitators (eg, enjoyment, parental and peer support and positive reinforcement) [47,90,91].

PtC 7: rheumatologists and HPRs should consider offering a variety of physical activity formats that align with young people's preferences and disease requirements

Although the WHO guidelines emphasise quantitative aspects of physical activity (frequency, duration and intensity), the task force underscores the importance of considering qualitative factors (ie, contextual elements) to optimise enjoyment [92], improve self-esteem [93] and to achieve a better adherence to physical activity in young people with jRMDs. Thus, rheumatologists, HPRs and any other caregivers (eg, families, teachers) may consider contextual factors when offering physical activity for young people with jRMDs, including but not limited to, the type of activity (eg, aerobic activity, bone-strengthening and muscle-strengthening exercise, mind-body activities and sports), level of supervision (ie, supervised or unsupervised), delivery mode (eg, face to face or remote), physical environment (ie, schools, community settings or sports clubs; indoors vs outdoors), social environment (eg, individual sports or team sports) and equipment (eg, ropes, balls).

The task force highlights the need for specialised assistance in offering variations in contextual factors of physical activity to reduce monotony, while ensuring the activity is safe and well-tolerated [90]. Patient education for physical activity may help young people with jRMDs to self-regulate physical activity levels

according to their capabilities and daily symptoms fluctuations. Importantly, engaging in a variety of physical activity may help them discover their personal preferences.

DISCUSSION

This manuscript presents the EULAR PtC for patient education in physical activity and self-management of pain in young people with jRMDs during transitional care, establishing two OAPs and seven PtC. These PtC are the consensus of a diverse, multidisciplinary and international panel of 26 members including paediatric and adult rheumatologists, HPRs and people with RMDs. The discussions of this task force were informed by a broad review of the available literature (including previous guidelines [35–38]) and semistructured individual interviews with each task force member. Most PtC were based on weak evidence (ie, expert opinion) due to the limited evidence available in the literature. PtC 5 and PtC 6, in contrast, were based on moderate evidence (ie, individual cohort study). Overall, task force agreement was high, except for PtC 4, which was lower due to challenges in identifying a practical method for measuring health literacy in daily clinical practice in transitional care.

The 2017 EULAR/PReS task force introduced general guidelines for transitional care in young people with jRMDs. Building on the progress initiated by those general guidelines, our task force viewed it as a next step to provide specific guidance on patient education in physical activity and self-management of pain during transitional care in rheumatology. We identified these areas as particularly urgent due to the high prevalence of physical inactivity and poor self-management of pain among young people with jRMDs, which may negatively impact their long-term outcomes and health-related quality of life. Thus, addressing these two key issues offers a valuable step forward in promoting continuity of care in rheumatology. To further advance transitional care in rheumatology, the task force outlined a research agenda emphasising the importance of experimentally evaluating the effects of existing transitional care programmes and, if needed, optimise them or develop new programmes (box 1).

Given the urgent need to raise awareness about the importance of ensuring continuity of care in rheumatology, our immediate goal is to promote the present PtC. We plan to disseminate these guidelines to both academic and layperson audiences through various channels, including scientific manuscripts, presentations and leaflets, as well as participation in mass media such as interviews, podcasts and social media campaigns. To maximise their reach and visibility, these materials will be translated into as many languages as possible.

Considering the limited availability and access to transitional care programmes of excellence in rheumatology, promoting the broad implementation of these PtC is a key priority for our task force. While digital health solutions (eg, apps) offer potential for patient education in physical activity and self-management of pain, they must be approached with caution due to the need for individualised strategies and limited supporting evidence. Additionally, placing the burden on patients is always problematic and particularly without adopting system-wide changes to support them. For instance, we highlight the need for caregivers training and ensuring that educational resources are both accessible and affordable.

A limitation of these PtC is that the task force decided to focus on patient education in physical activity and self-management of pain, thereby excluding other important areas, such as patient education in self-management of fatigue. Expanding the scope to include additional outcomes would have been excessively ambitious and could have compromised the feasibility of the task force's work. Additionally, most of the available literature is from studies conducted in relatively homogeneous populations; predominantly, white individuals with Juvenile Idiopathic Arthritis from developed countries, with scarce attention to individuals from minorities, populations at risk of social exclusion and those underrepresented in research or with limited access to care. In addition to raising questions about generalisability to other populations, this may be particularly relevant during transitional care as diversity continues to increase in younger populations. While these limitations should be acknowledged, the present PtC emphasise the importance of personalised transitional care tailored to individual needs, priorities, preferences, capabilities, resources and contexts. In addition, a potential limitation of the consensus process is the potential for bias in expert selection, which may lead to an overrepresentation of certain viewpoints, limiting the diversity of perspectives.

In conclusion, this EULAR PtC for patient education in physical activity and self-management of pain in young people with jRMDs during transitional care provide guidance to rheumatologists, HPRs, young people with jRMDs, families, any other caregivers (eg, teachers) and organisations. Our ultimate goal is to promote continuity of care in rheumatology, empowering young people with jRMDs to manage their health proactively and improve their long-term health outcomes. These PtC are intended to fill gaps in current practice, drive future research and potentially shape policies to ensure more standardised, person-centred transitional care across Europe.

Box 1 Research agenda

- To conduct high-quality studies that generate evidence supporting or refuting the present points to consider for patient education in physical activity and self-management of pain in transitional care in rheumatology, which is primarily based on expert consensus.
- To analyse the effects of existing programmes in these areas through well-designed randomised controlled trials that assess key outcomes in transitional care (eg, readiness to transfer, continuity of care) and, if needed, optimise them or develop new programmes.
- To develop sustainable, customised educational interventions in these areas, with developmentally appropriate content that aligns with the evolving priorities of young people's with juvenile-onset rheumatic and musculoskeletal diseases (jRMDs) as they grow.
- To explore the barriers and facilitators to implementing patient education in these areas into the daily lives of young people with jRMDs, including school and leisure time.
- To examine how digital health tools may support young people with jRMDs—and their families, healthcare professionals and any other caregivers (eg, teachers)—as they transition from paediatric to adult care, while promoting responsible use of technology and addressing issues such as time management and avoiding sedentary behaviour.

Contributors

All authors are members of the EULAR's task force HPR051. JC-I and RP-M were the fellows. FE-L was the convenor (guarantor). LC and FE-L were the co-methodologists. All authors

contributed to the work, read and finally approved the manuscript for submission.

Funding

This study was funded by EULAR (HPR051). RP-M and FE-L were supported by the Spanish Ministry of Science and Innovation (JDC2023-052717-I and RYC2021-034311-I, respectively).

Competing interests

JC-I: none declared, RP-M: none declared, EB-V: none declared, PA-V: none declared, SA: none declared, JA: none declared, IB: none declared, DC: none declared, MC: none declared, WC: none declared, DDC: none declared, AD: none declared, LL: payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Pfizer; AM: none declared, KM: grant/research support from Pfizer, Novartis and Medac, AFM: none declared, AN: none declared, SO: none declared, GP: none declared, ZS: none declared, TV: none declared, NMW: none declared, AZC: none declared, YP: none declared, LC: none declared, FE-L: none declared.

Patient and public involvement statement

Patients as research partners were involved in the design, conduction, reporting and dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication

Not applicable.

Data availability statement

Data sharing is not applicable as no datasets were generated for this study. All data relevant to the study are included in the article or uploaded as online supplemental information.

Orcid

Javier Courel-Ibáñez: http://orcid.org/0000-0003-2446-1875
Rafael Prieto-Moreno: http://orcid.org/0000-0002-7301-5831
Jordi Anton: http://orcid.org/0000-0002-8792-4219
Andrea Domján: http://orcid.org/0009-0003-3676-9911
Andréa Marques: http://orcid.org/0000-0002-2026-9926
Kirsten Minden: http://orcid.org/0000-0003-2775-0111
Aurélie Najm: http://orcid.org/0000-0002-6008-503X
Seza Ozen: http://orcid.org/0000-0003-2883-7868
Tomas Vetrovsky: http://orcid.org/0000-0003-2529-7069
Nico M. Wulffraat: http://orcid.org/0000-0001-9548-5562
Andrea Zacarias Crovato: http://orcid.org/0000-0002-1865-8850
Yeliz Prior: http://orcid.org/0000-0001-9831-6254
Loreto Carmona: http://orcid.org/0000-0002-4401-2551
Fernando Estévez-López: http://orcid.org/0000-0003-2960-4142

REFERENCES

[1] Clemente D, Leon L, Foster H, et al. Transitional care for rheumatic conditions in Europe: current clinical practice and available resources. Pediatr Rheumatol 2017;15. doi: 10.1186/s12969-017-0179-8.
[2] Bohr A-H, Nielsen S, Müller K, et al. Reduced physical activity in children and adolescents with Juvenile Idiopathic Arthritis despite satisfactory control of inflammation. Pediatr Rheumatol 2015;13. doi: 10.1186/s12969-015-0053-5.
[3] Saidi O, Rochette E, Bourdier P, et al. Sleep in children and adolescents with juvenile idiopathic arthritis: a systematic review and meta-analysis of case-control studies. Sleep 2022;45. doi: 10.1093/sleep/zsab233.
[4] Schlichtiger J, Haas J-P, Barth S, et al. Education and employment in patients with juvenile idiopathic arthritis – a standardized comparison to the German general population. Pediatr Rheumatol 2017;15. doi: 10.1186/s12969-017-0172-2.
[5] Eleftheriou D, Isenberg DA, Wedderburn LR, et al. The coming of age of adolescent rheumatology. Nat Rev Rheumatol 2014;10:187–93. doi: 10.1038/nrrheum.2013.202.
[6] Oliveira Ramos F, Rodrigues A, Magalhaes Martins F, et al. Health-related quality of life and disability in adults with juvenile idiopathic arthritis: comparison with adult-onset rheumatic diseases. RMD Open 2021;7:e001766. doi: 10.1136/rmdopen-2021-001766.
[7] Shoop-Worrall SJW, Kearsley-Fleet L, Thomson W, et al. How common is remission in juvenile idiopathic arthritis: A systematic review. Semin Arthritis Rheum 2017;47:331–7. doi: 10.1016/j.semarthrit.2017.05.007.
[8] Clemente D, Leon L, Foster H, et al. Systematic review and critical appraisal of transitional care programmes in rheumatology. Semin Arthritis Rheum 2016;46:372–9. doi: 10.1016/j.semarthrit.2016.06.003.
[9] Oliveira Ramos F, Zinterl C, Fonseca JE. A lifelong journey: Long-term perspectives on Juvenile Idiopathic Arthritis. Best Pract Res Clin Rheumatol 2024;38:101984. doi: 10.1016/j.berh.2024.101984.
[10] Hazel E, Zhang X, Duffy CM, et al. High rates of unsuccessful transfer to adult care among young adults with juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2010;8:1–6 2 10.1186/1546-0096-8-2.
[11] Jensen PT, Karnes J, Jones K, et al. Quantitative evaluation of a pediatric rheumatology transition program. Pediatr Rheumatol Online J 2015;13:17. doi: 10.1186/s12969-015-0013-0.
[12] Hersh AO, Pang S, Curran ML, et al. The challenges of transferring chronic illness patients to adult care: reflections from pediatric and adult rheumatology at a US academic center. Pediatr Rheumatol 2009;7:1–7. doi: 10.1186/1546-0096-7-13.
[13] Foster HE, Minden K, Clemente D, et al. EULAR/PReS standards and recommendations for the transitional care of young people with juvenile-onset rheumatic diseases. Ann Rheum Dis 2017;76:639–46. doi: 10.1136/annrheumdis-2016-210112.
[14] Mikola K, Rebane K, Arnstad ED, et al. Transitioning patients with juvenile idiopathic arthritis to adult care: the Nordic experience. Pediatr Rheumatol 2022;20. doi: 10.1186/s12969-022-00742-2.
[15] Barnabe C, Chomistek K, Luca N, et al. National Priorities for High-quality Rheumatology Transition Care for Youth in Canada. J Rheumatol 2021;48:426–33. doi: 10.3899/jrheum.200790.
[16] Berben L, Sigg N, Daly ML, et al. Current practice of transitional care for adolescents and young adults in Swiss paediatric and adult rheumatology centres. Swiss Med Wkly 2021;151. doi: 10.4414/smw.2021.w30046.
[17] Valenzuela PL, Ruilope LM, Santos-Lozano A, et al. Exercise benefits in cardiovascular diseases: from mechanisms to clinical implementation. Eur Heart J 2023;44:1874–89. doi: 10.1093/eurheartj/ehad170.
[18] Metsios GS, Stavropoulos-Kalinoglou A, Kitas GD. The role of exercise in the management of rheumatoid arthritis. Expert Rev Clin Immunol 2015;11:1121–30. doi: 10.1586/1744666X.2015.1067606.
[19] Campbell KL, Winters-Stone KM, Wiskemann J, et al. Exercise Guidelines for Cancer Survivors: Consensus Statement from International Multidisciplinary Roundtable. Med Sci Sports Exerc 2019;51:2375–90. doi: 10.1249/MSS.0000000000002116.
[20] Wang F, Boros S. The effect of physical activity on sleep quality: a systematic review. Eur J Physiother 2021;23:11–8. doi: 10.1080/21679169.2019.1623314.
[21] Rodriguez-Ayllon M, Cadenas-Sánchez C, Estévez-López F, et al. Role of physical activity and sedentary behavior in the mental health of preschoolers, children and adolescents: a systematic review and meta-analysis. Sports Med 2019;49:1383–410.
[22] Marquez DX, Aguiñaga S, Vásquez PM, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med 2020;10:1098–109. doi: 10.1093/tbm/ibz198.
[23] Sveaas SH, Smedslund G, Hagen KB, et al. Effect of cardiorespiratory and strength exercises on disease activity in patients with inflammatory

rheumatic diseases: a systematic review and meta-analysis. Br J Sports Med 2017;51:1065–72. doi: 10.1136/bjsports-2016-097149.
[24] Katz P, Andonian BJ, Huffman KM. Benefits and promotion of physical activity in rheumatoid arthritis. Curr Opin Rheumatol 2020;32:307–14. doi: 10.1097/BOR.0000000000000696.
[25] Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54:1451–62. doi: 10.1136/bjsports-2020-102955.
[26] Henderson CJ, Lovell DJ, Specker BL, et al. Physical activity in children with juvenile rheumatoid arthritis: Quantification and evaluation. Arthritis & Rheum 1995;8:114–9. doi: 10.1002/art.1790080210.
[27] Kimura Y, Walco GA. Treatment of chronic pain in pediatric rheumatic disease. Nat Clin Pract Rheumatol 2007;3:210–8. doi: 10.1038/ncprheum0458.
[28] Nordal E, Rypdal V, Arnstad ED, et al. Participation in school and physical education in juvenile idiopathic arthritis in a Nordic long-term cohort study. Pediatr Rheumatol Online J 2019;17:44. doi: 10.1186/s12969-019-0341-6.
[29] Schanberg LE, Anthony KK, Gil KM, et al. Daily pain and symptoms in children with polyarticular arthritis. Arthritis & Rheum 2003;48:1390–7. doi: 10.1002/art.10986.
[30] Lim LSH, Ekuma O, Marrie RA, et al. A Population-based Study of Grade 12 Academic Performance in Adolescents With Childhood-onset Chronic Rheumatic Diseases. J Rheumatol 2022;49:299–306. doi: 10.3899/jrheum.201514.
[31] Kembe J, Regardt M. Experiences of participation in activities among girls with juvenile idiopathic arthritis: A qualitative study. J Child Health Care 2023;27:599–611. doi: 10.1177/13674935221083167.
[32] Packham JC, Hall MA. Long-term follow-up of 246 adults with juvenile idiopathic arthritis: education and employment. Rheumatology (Oxford) 2002;41:1436–9. doi: 10.1093/rheumatology/41.12.1436.
[33] McDonagh JE, Farre A. Transitional Care in Rheumatology: a Review of the Literature from the Past 5 Years. Curr Rheumatol Rep 2019;21:57. doi: 10.1007/s11926-019-0855-4.
[34] van der Heijde D, Aletaha D, Carmona L, et al. 2014 Update of the EULAR standardised operating procedures for EULAR-endorsed recommendations. Ann Rheum Dis 2015;74:8–13. doi: 10.1136/annrheumdis-2014-206350.
[35] Zangi HA, Ndosi M, Adams J, et al. EULAR recommendations for patient education for people with inflammatory arthritis. Ann Rheum Dis 2015;74:954–62. doi: 10.1136/annrheumdis-2014-206807.
[36] Geenen R, Overman CL, Christensen R, et al. EULAR recommendations for the health professional's approach to pain management in inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018;77:797–807. doi: 10.1136/annrheumdis-2017-212662.
[37] Rausch Osthoff A-K, Niedermann K, Braun J, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018;77:1251–60. doi: 10.1136/annrheumdis-2018-213585.
[38] Nikiphorou E, Santos EJF, Marques A, et al. 2021 EULAR recommendations for the implementation of self-management strategies in patients with inflammatory arthritis. Ann Rheum Dis 2021;80:1278–85. doi: 10.1136/annrheumdis-2021-220249.
[39] Oxford Centre for Evidence-Based Medicine. Levels of evidence (Updated in 2012) Available: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009; 2009.
[40] Šulinskaitė K, Zagurskienė D, Blaževičienė A. Patients' health literacy and health behaviour assessment in primary health care: evidence from a cross-sectional survey. BMC Prim Care 2022;23:223. doi: 10.1186/s12875-022-01809-5.
[41] Şenol Y, Şenol FB. Health Promotion in Preschool Children. Children (Basel) 2023;10:1385. doi: 10.3390/children10081385.
[42] Huckerby L, McDonagh JE, Lee RR. Growing up with chronic pain: a systematic review of the evidence in transitional care for young people with chronic pain. Rheumatol Adv Pract 2023;7:i19–25. doi: 10.1093/rap/rkad006.
[43] Ross E, Abulaban K, Kessler E, et al. Non-pharmacologic therapies in treatment of childhood-onset systemic lupus erythematosus: A systematic review. Lupus (Los Angel) 2022;31:864–79. doi: 10.1177/09612033221094704.
[44] Lalloo C, Harris LR, Hundert AS, et al. The iCanCope pain self-management application for adolescents with juvenile idiopathic arthritis: a pilot randomized controlled trial. Rheumatology (Sunnyvale) 2021;60:196–206. doi: 10.1093/rheumatology/keaa178.
[45] Patton M, Martin-Feeney DS, Allemang B, et al. What skills do adolescents and young adults desire as they prepare for adult health care? H C T 2024;2:100049. doi: 10.1016/j.hctj.2024.100049.
[46] Lunde CE, Fisher E, Donovan E, et al. Cutting the cord? Parenting emerging adults with chronic pain. Pediatr Neonat Pain 2022;4:136–47. doi: 10.1002/pne2.12072.
[47] Gualano B, Bonfa E, Pereira RMR, et al. Physical activity for paediatric rheumatic diseases: standing up against old paradigms. Nat Rev Rheumatol 2017;13:368–79. doi: 10.1038/nrrheum.2017.75.
[48] Caes L, Dick B, Duncan C, et al. The Cyclical Relation Between Chronic Pain, Executive Functioning, Emotional Regulation, and Self-Management. J Pediatr Psychol 2021;46:286–92. doi: 10.1093/jpepsy/jsaa114.
[49] Jones A, Caes L, Eccleston C, et al. The sands of time: Adolescents' temporal perceptions of peer relationships and autonomy in the context of living with chronic pain. Pediatr Neonat Pain 2022;4:110–24. doi: 10.1002/pne2.12071.
[50] Kashikar-Zuck S, Black WR, Pfeiffer M, et al. Pilot Randomized Trial of Integrated Cognitive-Behavioral Therapy and Neuromuscular Training for Juvenile Fibromyalgia: The FIT Teens Program. J Pain 2018;19:1049–62. doi: 10.1016/j.jpain.2018.04.003.
[51] McColl J, Semalulu T, Beattie KA, et al. Transition Readiness in Adolescents With Juvenile Idiopathic Arthritis and Childhood-Onset Systemic Lupus Erythematosus. ACR Open Rheumatol 2021;3:260–5. doi: 10.1002/acr2.11237.
[52] Rapley T, Farre A, Parr JR, et al. Can we normalise developmentally appropriate health care for young people in UK hospital settings? An ethnographic study. BMJ Open 2019;9:e029107. doi: 10.1136/bmjopen-2019-029107.
[53] DeCoste C, Moaf P, Mohamed I, et al. Adolescent Health Care Needs and Relationship to Disease in Patients With Childhood–Onset Systemic Lupus Erythematosus. Arthritis Care Res (Hoboken) 2024;76:841–9. doi: 10.1002/acr.25297.
[54] Leon L, Clemente D, Heredia C, et al. Self-esteem, self-concept, and body image of young people with rheumatic and musculoskeletal diseases: A systematic literature review. Semin Arthritis Rheum 2024;68:152486. doi: 10.1016/j.semarthrit.2024.152486.
[55] McDonagh JE, Tattersall R, Clinch J, et al. Developmentally appropriate transitional care during the Covid-19 pandemic for young people with juvenile-onset rheumatic and musculoskeletal diseases: the rationale for a position statement. Pediatr Rheumatol Online J 2021;19:136. doi: 10.1186/s12969-021-00609-y.
[56] AlMarshedi A, Wills G, Ranchhod A. Gamifying Self-Management of Chronic Illnesses: A Mixed-Methods Study. JMIR Serious Games 2016;4:e14. doi: 10.2196/games.5943.
[57] Stutvoet MD, Levelt L, Hrehovcsik MM, et al. Gamification in eHealth for Chronic Disease Self-Management in Youth: A Systematic Review. Games Health J 2024;13:314–31. doi: 10.1089/g4h.2023.0111.
[58] Lanis A, Tu E, Peskin M, et al. Storytelling of Young Adults with Chronic Rheumatologic Illnesses: A Pilot Study. Healthcare (Basel) 2022;10:1979. doi: 10.3390/healthcare10101979.
[59] Stinson J, Ahola Kohut S, Forgeron P, et al. The iPeer2Peer Program: a pilot randomized controlled trial in adolescents with Juvenile Idiopathic Arthritis. Pediatr Rheumatol Online J 2016;14:48. doi: 10.1186/s12969-016-0108-2.
[60] Cavallo S, April KT, Duffy C, et al. Development and preliminary acceptability of jiactiv, a social media-based program promoting engagement in physical activity among young people living with juvenile idiopathic arthritis. Arthritis Rheumatol 2021;73:540–1.
[61] Arman N, Tarakci E, Tarakci D, et al. Effects of Video Games-Based Task-Oriented Activity Training (Xbox 360 Kinect) on Activity Performance and Participation in Patients With Juvenile Idiopathic Arthritis: A Randomized Clinical Trial. Am J Phys Med Rehabil 2019;98:174–81. doi: 10.1097/PHM.0000000000001001.
[62] Lazaroff SM, Meara A, Tompkins MK, et al. How Do Health Literacy, Numeric Competencies, and Patient Activation Relate to Transition Readiness in Adolescents and Young Adults With Rheumatic Diseases? Arthritis Care Res (Hoboken) 2019;71:1264–9. doi: 10.1002/acr.23739.
[63] Bray L, Wilkinson C, Bruce C, et al. "It's my back…"; developing the coming to spinal clinic resource to improve the health literacy of young people with adolescent idiopathic scoliosis and their parents. J Child Health Care 2023;27:628–42. doi: 10.1177/13674935221083767.
[64] Doğan Y, Karaca NB, Buran S, et al. Health Literacy Levels of Patients With Juvenile Idiopathic Arthritis and Their Parents. Health Expect 2024;27:e14117. doi: 10.1111/hex.14117.
[65] Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health 2023;5:1264780. doi: 10.3389/fdgth.2023.1264780.
[66] Gazmararian JA, Williams MV, Peel J, et al. Health literacy and knowledge of chronic disease. Patient Educ Couns 2003;51:267–75. doi: 10.1016/s0738-3991(02)00239-2.
[67] Arlinghaus KR, Johnston CA. Advocating for Behavior Change With Education. Am J Lifestyle Med 2018;12:113–6. doi: 10.1177/1559827617745479.
[68] Veshovda S, Eik H, Andersen MH, et al. Health literacy and musculoskeletal disorders in adolescents: a scoping review. BMJ Open 2023;13:e072753. doi: 10.1136/bmjopen-2023-072753.

[69] Araújo-Soares V, Hankonen N, Presseau J, et al. Developing Behavior Change Interventions for Self-Management in Chronic Illness: An Integrative Overview. Eur Psychol 2019;24:7–25. doi: 10.1027/1016-9040/a000330.
[70] Arafa A, Yasui Y, Kokubo Y, et al. Lifestyle Behaviors of Childhood and Adolescence: Contributing Factors, Health Consequences, and Potential Interventions. Am J Lifestyle Med 2024. doi: 10.1177/15598276241245941.
[71] Souilla L, Larsen AC, Juhl CB, et al. Childhood and adolescence physical activity and multimorbidity later in life: A systematic review. J Multimorbidity Comorbidity 2024;14. doi: 10.1177/26335565241231403.
[72] Zare N, Mansoubi M, Coe S, et al. An investigation into the relationship between nutritional status, dietary intake, symptoms and health-related quality of life in children and young people with juvenile idiopathic arthritis: a systematic review and meta-analysis. BMC Pediatr 2023;23:3. doi: 10.1186/s12887-022-03810-4.
[73] Yarema N, Fedortsiv O, Palasiuk B. Influence of ω-3 polyunsaturated fatty acids on the structure of immunocompetent cell membranes and the parameters of cellular and humoral immunity in children with rheumatoid arthritis. fmpcr 2018;20:78–82. doi: 10.5114/fmpcr.2018.73707.
[74] Dilandro G, De Cosmi V, Corona F, et al. May biscuits contribute to iron balance? An observation in children with juvenile idiopatic arthritis. Int J Food Sci Nutr 2015;66:811–4. doi: 10.3109/09637486.2015.1077791.
[75] Stark LJ, Janicke DM, McGrath AM, et al. Prevention of osteoporosis: a randomized clinical trial to increase calcium intake in children with juvenile rheumatoid arthritis. J Pediatr Psychol 2005;30:377–86. doi: 10.1093/jpepsy/jsi061.
[76] Kalkwarf HJ, Khoury JC, Lanphear BP. Milk intake during childhood and adolescence, adult bone density, and osteoporotic fractures in US women. Am J Clin Nutr 2003;77:257–65. doi: 10.1093/ajcn/77.1.257.
[77] Surís J-C, Michaud P-A, Akre C, et al. Health risk behaviors in adolescents with chronic conditions. Pediatrics 2008;122:e1113–8. doi: 10.1542/peds.2008-1479.
[78] Wisk LE, Weitzman ER. Substance use patterns through early adulthood: Results for youth with and without chronic conditions. Am J Prev Med 2016;51:33–45. doi: 10.1016/j.amepre.2016.01.029.
[79] Weitzman ER, Minegishi M, Dedeoglu F, et al. Disease-Tailored Brief Intervention for Alcohol Use Among Youths With Chronic Medical Conditions. JAMA Netw Open 2024;7:e2419858. doi: 10.1001/jamanetworkopen.2024.19858.
[80] Albinni B, de Zambotti M, Iacovides S, et al. The complexities of the sleep-pain relationship in adolescents: A critical review. Sleep Med Rev 2023;67:101715. doi: 10.1016/j.smrv.2022.101715.
[81] Engelen MM, Knoll JL, Rabsztyn PRI, et al. Sexual Health Communication Between Healthcare Professionals and Adolescents with Chronic Conditions in Western Countries: An Integrative Review. Sex Disabil 2020;38:191–216. doi: 10.1007/s11195-019-09597-0.
[82] Savel C, Chausset A, Berland P, et al. Survey of adolescents' needs and parents' views on sexual health in juvenile idiopathic arthritis. Pediatr Rheumatol 2023;21. doi: 10.1186/s12969-023-00884-x.
[83] Witczak BN, Bollerslev J, Godang K, et al. Body composition in long-standing juvenile dermatomyositis: associations with disease activity, muscle strength and cardiometabolic measures. Rheumatology (Sunnyvale) 2022;61:2959–68. doi: 10.1093/rheumatology/keab805.
[84] Yang S, Luo C, Zheng W, et al. Patterns of body composition and alteration after treatment in patients with newly diagnosed idiopathic inflammatory myopathies. Rheumatology (Oxford) 2022;62:270–80. doi: 10.1093/rheumatology/keac286.
[85] Jiang I, Major G, Singh-Grewal D, et al. Patient and parent perspectives on transition from paediatric to adult healthcare in rheumatic diseases: an interview study. BMJ Open 2021;11:e039670. doi: 10.1136/bmjopen-2020-039670.
[86] Gwinnutt JM, Wieczorek M, Balanescu A, et al. 2021 EULAR recommendations regarding lifestyle behaviours and work participation to prevent progression of rheumatic and musculoskeletal diseases. Ann Rheum Dis 2023;82:48–56. doi: 10.1136/annrheumdis-2021-222020.
[87] Kuntze G, Nesbitt C, Whittaker JL, et al. Exercise Therapy in Juvenile Idiopathic Arthritis: A Systematic Review and Meta-Analysis. Arch Phys Med Rehabil 2018;99:178–93. doi: 10.1016/j.apmr.2017.05.030.
[88] Nascimento Leite M, Kamper SJ, O'Connell NE, et al. Physical activity and education about physical activity for chronic musculoskeletal pain in children and adolescents. Cochrane Database Syst Rev 2023;7:CD013527. doi: 10.1002/14651858.CD013527.pub2.
[89] Pandya J, Menino Rosenbluth LJ, Adams AB. Physical Activity and Sports for Children With Juvenile Idiopathic Arthritis. HSS J 2024;20:409–15. doi: 10.1177/15563316241247828.
[90] Iversen MD, Andre M, von Heideken J. Physical Activity Interventions in Children with Juvenile Idiopathic Arthritis: A Systematic Review of Randomized Controlled Trials. PHMT 2022;13:115–43. doi: 10.2147/PHMT.S282611.
[91] Brandelli YN, Stone M, Westheuser V, et al. Parent Risk Perceptions, Physical Literacy, and Fundamental Movement Skills in Children With Juvenile Idiopathic Arthritis. Pediatr Phys Ther 2022;34:536–44. doi: 10.1097/PEP.0000000000000948.
[92] Vella SA, Aidman E, Teychenne M, et al. Optimising the effects of physical activity on mental health and wellbeing: A joint consensus statement from Sports Medicine Australia and the Australian Psychological Society. J Sci Med Sport 2023;26:132–9. doi: 10.1016/j.jsams.2023.01.001.
[93] Burnet K, Kelsch E, Zieff G, et al. How fitting is F.I.T.T.?: A perspective on a transition from the sole use of frequency, intensity, time, and type in exercise prescription. Physiol Behav 2019;199:33–4. doi: 10.1016/j.physbeh.2018.11.007.

Ann Rheum Dis 84 (2025) 538–546

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Axial spondyloarthritis

The Assessment of SpondyloArthritis International Society (ASAS) Consensus-Based Expert Definition of Difficult-to-Manage, including Treatment-Refractory, Axial Spondyloarthritis

Denis Poddubnyy[1,2,*], Victoria Navarro-Compán[3], Murat Torgutalp[2], Suzanne Arends[4], Sibel Zehra Aydin[5,6], Simone Battista[7], Filip van den Bosch[8], Christine Bundy[9], Alberto Cauli[10], Jo Davies[11], Maxime Dougados[12], Tuncay Duruöz[13], Bassel El-Zorkany[14], Warren Fong[15,16], Floris van Gaalen[17], Rodrigo Garcia-Salinas[18], Marco Garrido Cumbrera[19], Pál Géher[20], Lianne Gensler[21], Simeon Grazio[22], Feng Huang[23], Mitsumasa Kishimoto[24], Robert Landewé[25,26], Ying Ying Leung[27], Pedro M Machado[28,29], Helena Marzo-Ortega[30], Bhowmik Meghnathi[31,32], Anna Molto[33], Elena Nikiphorou[34], Sofia Ramiro[17], Martin Rudwaleit[35], Carla G S Saad[36], Alexandre Sepriano[17,37], James Wei[38,39,40], Xenofon Baraliakos[41], Désirée van der Heijde[42], on behalf of ASAS

[1] *Division of Rheumatology, Department of Medicine, University of Toronto and University Health Network, Toronto, ON, Canada*
[2] *Department of Gastroenterology, Infectiology and Rheumatology (including Nutrition Medicine), Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany*
[3] *Department of Rheumatology, La Paz University Hospital, IdiPaz, Madrid, Spain*
[4] *Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*
[5] *Division of Rheumatology, Department of Medicine, University of Ottawa, Ottawa, ON, Canada*
[6] *Ottawa Hospital Research Institute, Ottawa, ON, Canada*
[7] *Centre for Human Movement and Rehabilitation, School of Health and Society, University of Salford, Salford, Greater Manchester, UK*
[8] *Department of Rheumatology, Ghent University, Ghent University Hospital, Ghent, Belgium*
[9] *School of Healthcare Sciences, Cardiff University, Wales, UK*
[10] *Rheumatology Unit, Department of Medicine and Public Health, AOU and University of Cagliari, Cagliari, Italy*
[11] *Axial Spondyloarthritis International Federation (ASIF), London, UK*
[12] *Department of Rheumatology, Hôpital Cochin, University Paris Cité, Paris, France*
[13] *PMR Department, Rheumatology Division, Marmara University School of Medicine, Istanbul, Turkey*
[14] *Department of Rheumatology, Cairo University, Cairo, Egypt*
[15] *Department of Rheumatology and Immunology, Singapore General Hospital, Singapore*
[16] *Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
[17] *Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands*

*Correspondence to Dr Denis Poddubnyy, Division of Rheumatology, Department of Medicine, University of Toronto, Toronto, ON, Canada.
E-mail address: denis.poddubnyy@uhn.ca (D. Poddubnyy).
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.01.035

[18] *Rheumatology Unit, Hospital Italiano de La Plata – Universidad Nacional de La Plata, La Plata, Argentina*
[19] *Health & Territory Research (HTR), Universidad de Sevilla, Seville, Spain*
[20] *Semmelweis University, Budapest, Hungary*
[21] *Division of Rheumatology, University of California San Francisco, San Francisco, CA, USA*
[22] *Department of Rheumatology, Physical and Rehabilitation Medicine, Clinical University Centre, Sestre Milosrdnice, Zagreb, Croatia*
[23] *Department of Rheumatology and Immunology, Chinese PLA General Hospital, Beijing, China*
[24] *Department of Nephrology and Rheumatology, Kyorin University School of Medicine, Tokyo, Japan*
[25] *Department of Rheumatology and Clinical Immunology, Amsterdam University Medical Center, Amsterdam, The Netherlands*
[26] *Department of Rheumatology, Zuyderland Medical Center, Heerlen, The Netherlands*
[27] *Department of Rheumatology & Immunology, Singapore General Hospital, Duke-NUS Medical School, Singapore*
[28] *Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London, UK*
[29] *Department of Rheumatology, Division of Medicine, University College London, London, UK*
[30] *NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust and Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK*
[31] *Department of Rheumatology & Immunology, SVP Hospital & Smt. NHL Municipal Medical College, Ahmedabad, India*
[32] *Department of Rheumatology, Marengo CIMS Hospital & RheumaCARE, Ahmedabad, India*
[33] *Rheumatology Department Cochin Hospital, APHP, INSERM U-1153, Centre de Recherche en Epidémiologie et Statistiques, Université Paris Cité, Paris, France*
[34] *Rheumatology Department, King's College Hospital, Centre for Rheumatic Diseases, Centre for Education, King's College London, London, UK*
[35] *Department of Internal Medicine and Rheumatology, Klinikum Bielefeld, University of Bielefeld, Bielefeld, Germany*
[36] *Rheumalogy Division, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil*
[37] *NOVA Medical School, UNL, Lisbon, Portugal*
[38] *Department of Allergy, Immunology & Rheumatology, Chung Shan Medical University Hospital, Taichung, Taiwan*
[39] *Department of Nursing, Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan*
[40] *Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan*
[41] *Rheumazentrum Ruhrgebiet, Herne and Ruhr-University, Bochum, Germany*
[42] *Department of Rheumatology, Leiden University, Leiden, The Netherlands*

ARTICLE INFO

Article history:
Received 18 September 2024
Received in revised form 15 January 2025
Accepted 16 January 2025

ABSTRACT

Objectives: To develop a consensus-based expert definition of difficult-to-manage (D2M) axial spondyloarthritis (axSpA), incorporating treatment-refractory (TR) disease.
Methods: A literature review was conducted in 2022 to identify potential definitions for D2M/TR axSpA from prior studies, followed by a 2-round Delphi consensus process conducted in 2022 and 2023 to identify components of D2M axSpA. Based on the results of the Delphi process, a draft of the D2M axSpA definition was developed and presented to the expert task force, including patient representation, and, subsequently, to the Assessment of SpondyloArthritis International Society (ASAS) membership for endorsement in January 2024.
Results: Consensus was reached on a D2M definition encapsulating treatment failure (treatment according to the ASAS-European Alliance of Associations for Rheumatology recommendations and failure of ≥2 biological or targeted synthetic disease-modifying antirheumatic drugs with different mechanisms of action unless contraindicated), suboptimal disease control, and physician or patient acknowledgement of problematic signs/symptoms in patients diagnosed with axSpA by the rheumatologist. This definition represents a broad concept that includes various reasons that lead to an unsatisfactory treatment outcome. TR axSpA is covered by the D2M definition but requires a history of treatment failure, the presence of objective signs of inflammatory activity, and the exclusion of noninflammatory reasons for nonresponse. The proposed D2M definition incorporating TR disease was endorsed by ASAS at the annual meeting in January 2024, with 89% votes (109/123) in favour of it.
Conclusions: The ASAS D2M axSpA definition, including TR disease, allows for identifying patients with unmet needs, paving the way for further research in this condition and its clinical care improvement.

INTRODUCTION

Axial spondyloarthritis (axSpA) is an immune-mediated inflammatory condition primarily affecting the axial skeleton (spine and sacroiliac joints) [1,2]. Based on the presence or absence of definitive radiographic sacroiliitis according to the radiographic criterion of the modified New York criteria [3], axSpA can be classified as radiographic, historically known as ankylosing spondylitis (AS) [4], or nonradiographic, respectively.

The first-line therapy for axSpA consists of nonsteroidal anti-inflammatory drugs (NSAIDs) in conjunction with nonpharmacological interventions, including regular exercise and physiotherapy [5]. In patients where first-line therapy is ineffective or

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Despite the availability of several efficacious treatment options, a significant proportion of patients with axial spondyloarthritis (axSpA) do not achieve satisfactory treatment outcomes.
- There has been no consistent or universally accepted definition of 'difficult-to-manage' (D2M) or 'treatment-refractory' (TR) axSpA, hindering the identification and management of patients with unmet needs.

WHAT THIS STUDY ADDS

- This study presents a consensus-based definition of D2M axSpA incorporating key elements such as treatment failure (defined by failure of at least 2 biological or targeted synthetic disease-modifying antirheumatic drugs), suboptimal disease control, and the perception of problematic signs and symptoms by both rheumatologists and patients.
- The definition includes a specific subset of patients with TR axSpA, characterised by objective signs of inflammatory activity despite optimal treatment, distinguishing them from patients who may have noninflammatory reasons for nonresponse.
- The definition was endorsed by the Assessment of SpondyloArthritis International Society, reflecting a broad consensus among experts in the field.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- The proposed definition provides a standardised framework for identifying patients with D2M and TR axSpA, facilitating targeted research into the underlying mechanisms, epidemiology, and potential interventions for these patient populations.
- This definition may inform policy-making, supporting the development of clinical guidelines and resource allocation for the management of patients with D2M axSpA.

not tolerated, second-line therapy should be considered. This includes biological or targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs). The first group of these drugs includes tumour necrosis factor (TNF) inhibitors and interleukin-17 (IL-17) inhibitors. The second group consists of Janus kinase (JAK) inhibitors [5].

The primary treatment target in axSpA is sustained remission, defined as the absence of clinical (signs and symptoms) and laboratory (primarily C-reactive protein [CRP]) indicators of disease activity [5,6]. The Axial Spondyloarthritis Disease Activity Score (ASDAS) is the preferred composite measure of disease activity in axSpA [5,6]; ASDAS < 1.3 is considered an inactive disease state and corresponds to remission, while ASDAS between 1.3 and 2.1 is classified as a low disease activity state and should be used as the alternative treatment target if remission is unachievable [7]. A clinically important improvement in ASDAS (Δ ASDAS ≥ 1.1) is used as a criterion for deciding on the continuation of b/tsDMARDs as outlined in the Assessment of SpondyloArthritis International Society (ASAS)-European Alliance of Associations for Rheumatology (EULAR) management recommendations for axSpA [5]. The ASAS 40 and 20 response criteria are frequently used in clinical trials as key outcome parameters [8].

Despite several efficacious anti-inflammatory treatment options, only about 40% to 50% of patients with axSpA achieve a relevant treatment response, and an even smaller proportion (approximately 10%-20%) reach remission or an inactive disease activity state within 16 to 24 weeks of treatment initiation, according to data from randomised controlled trials with b/tsDMARDs [9]. After the failure of 1 b/tsDMARD and in the presence of active disease, a switch to another b/tsDMARD is recommended [5]. However, a group of patients with non- or incomplete responses remains despite exposure to multiple advanced therapies. There are several potential reasons for nonresponse or partial response in axSpA, which may be related to the disease itself or factors other than inflammatory activity (eg, nonnociceptive pain mechanisms [10–12]). A misdiagnosis could also contribute to the observed nonresponse. The mechanisms underpinning nonresponse remain incompletely understood, and there are no evidence-based approaches to address this issue in clinical practice.

In recent years, the concept of 'difficult-to-treat' rheumatoid arthritis (RA) has evolved [13], leading to the development of specific recommendations for its management in clinical practice [14].

As part of the ASAS Difficult-to-Manage (D2M) axSpA initiative, which aims to define D2M axSpA and provide management guidance, we sought to develop a consensus-based expert definition of D2M axSpA, incorporating treatment-refractory (TR) disease, which will help facilitate further initiatives to improve the clinical care of these patients and research in this area.

METHODS

The process of developing the D2M definition involved forming a task force and conducting a literature review to inform the task force and a 2-round Delphi survey. Importantly, ASAS intended to develop a consensus-based definition and not classification criteria, which determined the methodology of the process. The first Delphi round was conducted among ASAS members to determine the main elements of the future definition. Following the discussion of the results within the task force and at the ASAS annual meeting, the second Delphi round focused on specific definition elements. Based on the results of this round, a draft of the D2M axSpA definition was developed and presented to the task force and, subsequently, to the ASAS membership for endorsement.

Task force

After the ASAS executive committee approved the project, a task force consisting of 29 rheumatologists and full ASAS members, 2 young ASAS representatives (a rheumatologist and an epidemiologist, full ASAS members), 2 patient representatives, 1 psychologist and behavioural medicine specialist, 1 physiotherapist, and 1 physical medicine specialist (full ASAS member) was convened.

Literature review

A literature review aimed at identifying potential definitions for D2M axSpA from prior studies was conducted in 2022, including a Medline (via PubMed) search using established review methods and the search terms outlined in Supplementary File S1. Following the RA model, the term 'difficult-to-treat' was utilised in the first search strategy. Recognising that the terminology in the literature might vary, a second, broader search strategy using terms reflecting treatment nonresponse in axSpA was implemented (Supplementary File S1). The search was performed for all types of articles published in English and based on studies in humans, with a publication date between 2012 and 2022. After excluding duplicates, titles and abstracts were

screened, followed by a full-text review by MT and DP. From the included publications, information was retrieved on the definition of treatment failure, the definition of active disease, and the terminology used to characterise the concept of 'difficult-to-manage' disease.

The first Delphi round

The first Delphi round focused on defining the following main elements of the definition: uncontrolled disease (clinical manifestations, composite outcome measures, objective signs of inflammation, and radiographic progression), treatment failure (types and number of treatment options applied), and other potential factors that might contribute to the D2M situation. At this stage, we used the term 'difficult-to-treat,' which was later replaced with 'difficult-to-manage' (see Results). The first Delphi round included 9 questions related to the D2M topic, including 1 open question (Supplementary File S2). It was conducted from November to December 2022, with all ASAS members and co-opted members of the D2M initiative—including 2 patient representatives, a psychologist and behavioural medicine specialist, and a physiotherapist, all of whom were non-ASAS members—invited to participate. The survey results were discussed with the members of the task force at a dedicated virtual meeting and subsequently with ASAS members at the annual meeting in January 2023.

The second Delphi round, draft definition, and endorsement

Taking the results of the discussions with task force members and at the ASAS meeting in January 2023 into account, we drafted the second Delphi round, which focused on refining the criteria for the D2M axSpA definition, including precise definitions for insufficient control of signs and symptoms, required treatment history, the number of prior b/tsDMARDs, discontinuations due to intolerability or side effects, and differentiation between primary and secondary treatment failure. This round included 11 questions related to the D2M definition, including an open question (Supplementary File S3). Before completing the survey, participants (the same group as in the first round) were informed about the outcomes of the previous stage. This round was conducted from September to October 2023 and was followed by discussions with the members of the task force and all ASAS members at the annual meeting in January 2024. As an outcome, the ASAS D2M definition was drafted, followed by a vote on endorsement by the full ASAS membership. A majority of votes in favour of the definition was sought for endorsement.

RESULTS

Literature review

A total of 198 publications were identified using both search strategies. After the exclusion of 12 duplicates, 186 publications were screened based on their titles and abstracts. A total of 134 publications were excluded: 128 were not related to the subject of interest, 4 were related to paediatrics, and 2 were not related to axSpA. Of the 52 publications whose full texts were evaluated, 41 were excluded as unrelated to the subject of interest. However, 4 new publications not captured by the original search strategies were included after a cross-reference check. Ultimately, 15 publications were included in the review: 2 case reports, 5 observational studies, 3 open-label clinical trials, 3 randomised controlled trials, and 2 review articles (see the Supplementary Fig and Supplementary Table). In summary, the literature review revealed only a few relevant works with no consistent definition of D2M axSpA due to the heterogeneity of the criteria used for defining active disease and history of treatment failure. Furthermore, there was no established terminology to characterise the group of interest.

The first Delphi round

A total of 212 ASAS members (both full and associate), along with 4 co-opted members of the D2M initiative, were invited; 123/212 (58%) responded and completed the survey in full. The majority of the respondents (53%) supported using an ASDAS ≥ 2.1 as an indicator of active disease in the context of D2M axSpA (referred to as difficult-to-treat in this round). Additionally, 73% indicated that objective signs of inflammatory activity (elevated CRP and/or active inflammation on magnetic resonance imaging [MRI]) should be incorporated into the definition of active disease. Moreover, 77% of the respondents believed that all manifestations of spondyloarthritis (axial, peripheral, and extramusculoskeletal) should be considered in the definition. Concerning the definition of treatment failure, the predominant response (46%) was '≥2 NSAIDs in full anti-inflammatory doses and ≥2 b/tsDMARDs with different modes of action,' without differentiating between primary and secondary nonresponse (79%). Regarding the question of whether intolerability or contraindications to NSAIDs or b/tsDMARDs should be considered as an alternative to insufficient efficacy in the definition, 48% of experts responded positively for both NSAIDs and b/tsDMARDs, while 16% supported this consideration for b/tsDMARDs only. According to 53% of the respondents, radiographic progression should be part of the definition, and 51% indicated that symptoms unrelated to the inflammatory activity of axSpA should not be considered.

In subsequent discussions with the task force and the ASAS membership during the ASAS 2024 annual meeting, it was decided to change the nomenclature from 'difficult-to-treat' to 'difficult-to-manage.' The main reason for this change is that the management of axSpA better incorporates all management aspects, not only drug treatment, and this is also in line with the ASAS-EULAR management recommendations. Furthermore, it was agreed that the D2M axSpA definition should be broad and inclusive, similar to the difficult-to-treat RA framework, as opposed to the TR scenario, which, being part of the D2M (and therefore covered by the D2M definition), relates to cases where inflammatory activity cannot be controlled with currently available treatments.

The second Delphi round

A total of 205 ASAS members (active at the time of invitation, both full and associate), along with 4 co-opted members, were invited, and 186/205 (91%) responded to the survey. In the first part of this round, we sought components for the definition of insufficient control of signs/symptoms of axSpA. The majority of respondents (59%) favoured using ASDAS ≥ 2.1 as the composite outcome measure threshold indicative of insufficient control of signs/symptoms in the context of the D2M axSpA definition. Other selected components included objective signs of inflammation (elevated CRP and active inflammation on MRI of sacroiliac joints or spine), which should be mandatory but only in TR patients (supported by 62% of the respondents), rapid radiographic spinal progression (defined as the development of >2 new syndesmophytes or bony bridges in 2 years [15], 63%),

and the presence of axSpA symptoms that cause a reduction in quality of life, even if axSpA is controlled according to the criteria mentioned above (80%).

The second part of the survey dealt with the treatment aspects of the D2M axSpA definition. The definition refers to the current version of the ASAS-EULAR management recommendations; therefore, no specific definition of minimal treatment duration was deemed necessary by 58% of the respondents. The leading response regarding the minimal sufficient treatment history (with 50% of the respondents in favour) was 'At least 2 b/tsDMARDs with different modes of action,' while 18% favoured 'At least 2 b/tsDMARDs with the same or different modes of action,' and 17% preferred 'At least 3 b/tsDMARDs with the same or different modes of action.' Treatment discontinuation due to intolerability/side effects was favoured by 75% of the respondents, and 70% favoured the incorporation of contraindications for treatment with b/tsDMARDs into the D2M definition, meaning that a patient could fulfil the definition without a trial of a b/tsDMARD. Furthermore, 51% of the respondents indicated no need for differentiation between primary and secondary nonresponse in the D2M context. Concurrently, 82% of the respondents believed that a lack of access to treatment should not be a part of the definition.

Draft definition and endorsement

The results of the second Delphi round were discussed by the task force. It was agreed to incorporate components of treatment history and insufficient symptom control into the draft definition based on the outcomes of the Delphi process. Specific attention was given to the items that received less than 70% of votes in the Delphi exercise. Additionally, the third component of the definition, which relates to the perception of the current situation as problematic by the rheumatologist and/or the patient, was also included following a discussion involving patient representatives. The decision to use 'and/or' instead of 'and' was made to ensure an appropriate representation of both patients' and physicians' views on the situation and to keep the definition inclusive.

The draft definition followed minor modifications to the proposed wording; consensus was reached. The final version of the ASAS D2M axSpA definition, as shown in Figure 1, was endorsed by ASAS at the annual meeting in January 2024 with 89% of the votes (109 out of 123 full members).

The D2M axSpA should only be applied to patients with a definite diagnosis of axSpA made by a rheumatologist. It consists of 3 main components: (1) treatment according to the ASAS-EULAR recommendations and failure of ≥2 b/tsDMARDs with different mechanisms of action (unless contraindicated); (2) insufficient control of signs/symptoms of axSpA; and (3) the present signs/symptoms being perceived as problematic by the rheumatologist and/or the patient (see Fig 1 for details).

TR axSpA, according to the endorsed definition, is considered a subgroup of D2M axSpA. Patients with D2M axSpA (assuming correct diagnosis, which should be the first step in the evaluation and treatment compliance) can be considered TR if ≥2 b/tsDMARDs failed, have high or very high disease activity according to ASDAS (ASDAS ≥ 2.1) plus objective signs of inflammatory activity (elevated CRP or active inflammation on MRI of sacroiliac joints or spine), and if other causes, likely responsible for signs and symptoms (concurrent conditions, noncompliance, etc) are excluded before making a decision on the presence of TR axSpA (Fig 2).

DISCUSSION

The developed expert consensus-based definition of D2M axSpA, including TR disease, is an important first step of the D2M initiative with the ultimate goal of improving treatment outcomes in axSpA. This initiative, led by ASAS, not only aims to define D2M and TR axSpA but also includes the development of management recommendations for D2M axSpA, encompassing TR cases. Here, we report the finalised definitions while the recommendation development process is ongoing. Importantly, patients are involved in the entire development process. Another important aspect is that the development process involved reaching a consensus among the members of an expert organisation; we did not aim to develop classification criteria, which

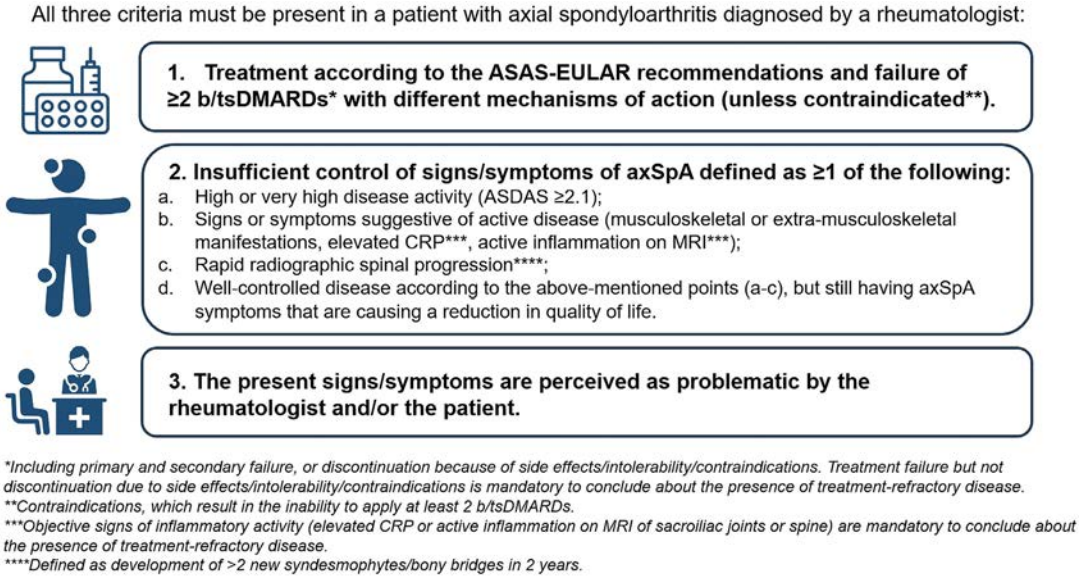


Figure 1. The Assessment of SpondyloArthritis International Society (ASAS) difficult-to-manage axial spondyloarthritis (axSpA) definition. ASDAS, Axial Spondyloarthritis Disease Activity Score; bDMARD, biologic disease-modifying antirheumatic drug; CRP, C-reactive protein; EULAR, European Alliance of Associations for Rheumatology; MRI, magnetic resonance imaging; tsDMARD, targeted synthetic disease-modifying antirheumatic drug.

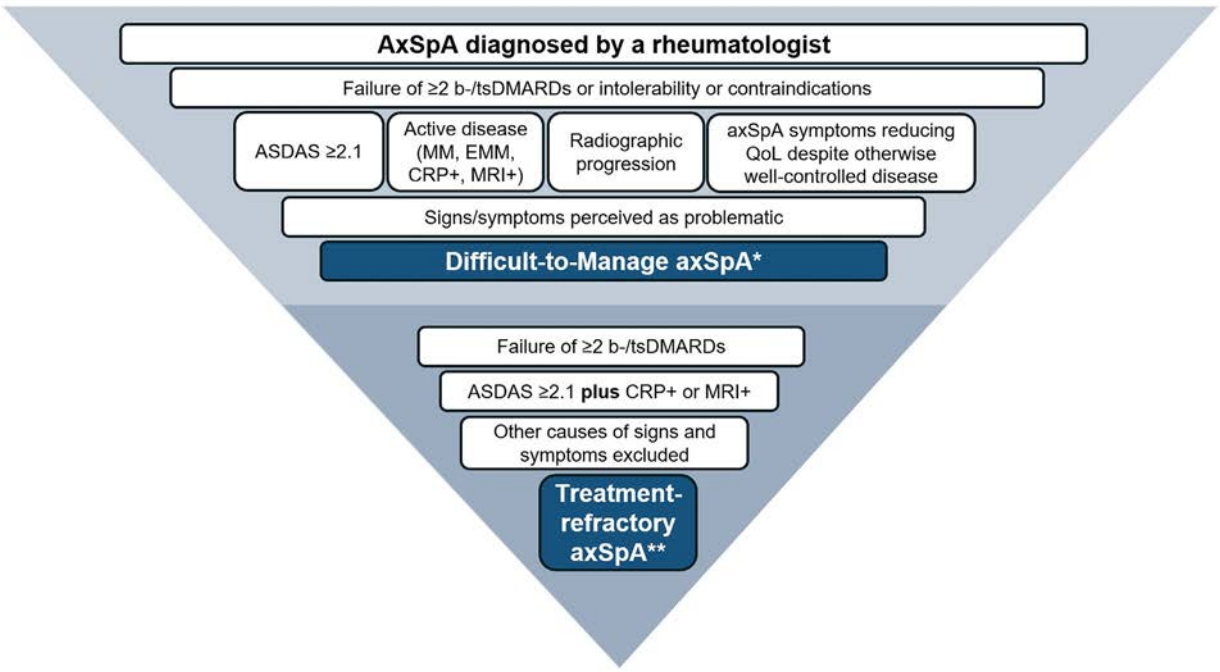


Figure 2. The difficult-to-manage axial spondyloarthritis (axSpA) construct. The starting point in this construct, which applies to both difficult-to-manage and treatment-refractory axSpA, is the diagnosis of axSpA by a rheumatologist. A difficult-to-manage situation (*) is present in cases of treatment failure (or intolerance/contraindications), indicators of uncontrolled signs/symptoms related to spondyloarthritis, and the perception of a problematic situation. A treatment-refractory situation (**) is present in patients meeting the definition of difficult-to-manage axSpA if there is evidence of treatment failure (assuming appropriate compliance, tolerance of the drugs, and sufficient treatment duration), high or very high disease activity according to the Axial Spondyloarthritis Disease Activity Score (ASDAS), and objective signs of uncontrolled inflammatory activity (as reflected by elevated C-reactive protein [CRP]: CRP + or inflammation on magnetic resonance imaging [MRI] of sacroiliac joints or spine: MRI +). It is assumed that other causes, likely responsible for signs and symptoms (including incorrect diagnosis, concurrent conditions, noncompliance, etc), are excluded before deciding on the presence of treatment-refractory axSpA. bDMARD, biologic disease-modifying antirheumatic drug; EMM, extramusculoskeletal manifestations; MM, musculoskeletal manifestations; QoL, quality of life; tsDMARD, targeted synthetic disease-modifying antirheumatic drug.

would have required a different methodological approach. We believe that a consensus-based expert approach is appropriate in this case, as there was no unified definition or terminology for the clinical situation described by the definition at the start of the initiative. Furthermore, we are defining not a disease or a permanent condition but rather a disease state that may change over time. Moreover, we followed a similar methodology that was used to develop the EULAR difficult-to-treat definition for RA [13].

The developed definition consists of 3 criteria, which must be present in a patient with axSpA diagnosed by a rheumatologist:

1. Treatment according to the ASAS-EULAR recommendations and failure of ≥2 b/tsDMARDs with different mechanisms of action (unless contraindicated).

This criterion defines the minimal requirement for treatment history and implies a lack of response to the standard treatment approach, including at least 2 b/tsDMARDs with different mechanisms of action. It implies the failure of b/tsDMARDs with proven efficacy in axSpA, which are incorporated in the ASAS-EULAR recommendations, currently TNF, IL-17, and JAK inhibitors. It is assumed that other treatment options, including NSAIDs and nonpharmacological measures, have been exhausted as well, either before or in parallel with b/tsDMARDs. In axSpA, only 3 classes of b/tsDMARDs are effective and approved for treatment; therefore, experts decided that at least 2 out of 3 classes should be tried before making a conclusion about the presence of D2M axSpA. Treatment failure includes both primary and secondary nonresponses since both may be associated with treatment challenges and a D2M situation. This criterion does not imply any specific time aspect: neither the duration of treatment, which must be in accordance with current recommendations, nor the timing of the treatment failure. However, this criterion should be considered in the context of other criteria; for instance, the definition of D2M axSpA will not be fulfilled in a patient with a history of b/tsDMARD secondary failure in the past if there are treatment options available or if the current treatment line is associated with a good clinical response. The same applies to the discontinuation of b/tsDMARDs due to intolerability or side effects (which are defined broadly in the D2M context as any event that results in the discontinuation of a drug). Patients with contraindications to 1 or several classes of b/tsDMARDs represent a particular group, which might be considered D2M despite the lack of formal evidence of b/tsDMARD failure. This means that this criterion may be fulfilled in a patient who failed 1 bDMARD and has contraindications to the use of others. When defining TR axSpA, which is a subgroup of D2M axSpA, evidence of treatment failure (at least 2 b/tsDMARDs belonging to different classes with proven efficacy in axSpA) and no discontinuation due to intolerability, side effects, or contraindications is required. This distinction is deemed necessary to differentiate between axSpA patients not responding to currently available treatment options and those who could have responded to the therapy but cannot receive it due to tolerability or safety issues.

2. Insufficient control of signs and symptoms of axSpA.

At least 1 of the following 4 indicators of insufficient control should be present: (i) high or very high disease activity according to the validated outcome measure ASDAS; (ii) presence of

active spondyloarthritis manifestations (musculoskeletal or extramusculoskeletal), including objective signs of inflammatory activity; (iii) rapid radiographic spinal progression, as defined by published data-driven criteria [15]; and (iv) other axSpA symptoms that are attributable to axSpA and cause a reduction in quality of life, even if points i to iv are not met. The criteria are broad and inclusive, aiming to capture the majority of clinical situations where axSpA might be considered insufficiently controlled. It is assumed that the mentioned signs and symptoms are present at, or are closely related temporally to, the time of the D2M status evaluation. The criterion of radiographic spinal progression specifically refers to the past 2 years. However, this must be contextualised with other parameters and the timing of treatment initiation since the effects of anti-inflammatory treatment, such as those shown for TNF inhibitors, typically become evident between years 2 and 4 of treatment [16]. There was significant discussion regarding the necessity of point iv. This point was retained in the final definition to ensure the inclusivity of the D2M definition, aiming to cover a broad range of clinical situations (eg, a patient with prominent fatigue or substantial functional limitations related to structural damage without significant pain or inflammatory activity). Although the mechanisms contributing may vary, inclusivity is vital for capturing these diverse scenarios. For defining TR disease, we propose that objective signs of inflammatory activity (elevated CRP that is attributable to axSpA and not to other causes, or active inflammation on MRI of sacroiliac joints or spine) be mandatory, in addition to the presence of high/very high disease activity according to ASDAS. Of note, patients experiencing rapid radiographic spinal progression would not be classified as TR if active inflammation is otherwise controlled. Radiographic progression in the spine can still be observed in the initial years following the introduction of effective anti-inflammatory treatment, often slowing over time [17]. Therefore, we do not classify patients with structural damage progression as TR if disease activity has been controlled by effective anti-inflammatory treatment, as there is a reasonable likelihood that progression will decelerate over time due to a time-shifted effect [16].

3. The present signs/symptoms are perceived as problematic by the rheumatologist and/or the patient.

This aspect is crucial as it brings together the physician's and patient's perspectives into the definition. It ensures that the evaluation of the D2M status is not merely based on formalised criteria relating to the number of previous treatment lines and composite outcome measures but also considers the global evaluation of the current situation in the context of the D2M concept. As mentioned above, in the broad and inclusive D2M definition, the opinion of the patient is as important as the opinion of the physician. No specific instruments are proposed to capture the perception of the disease as problematic from either the physician's or the patient's perspective.

What are the potential implications of the developed definition? We expect that it will stimulate research focusing on identifying reasons for D2M and would draw attention to D2M patients in daily clinical practice. We encourage investigators to prospectively collect information related to the elements of this definition in both interventional and observational studies on axSpA. The reasons for D2M may vary from setting to setting but most likely will belong to 1 of 2 main groups, which are important in both daily clinical practice and research contexts:

1. The true nonresponse to anti-inflammatory treatment resulted in a TR case. The exact frequency of this phenomenon, as well as the underlying mechanisms, warrant investigation, including the generation of epidemiological data, exploration of pathophysiology, and conduction of interventional studies.
2. Signs and symptoms not caused by inflammation but rather by nonnociceptive pain mechanisms (ie, nociplastic or neuropathic pain), concurrent conditions (which might be present even if the diagnosis of axSpA is correct), and other factors—including but not limited to socio-psychological aspects, work, beliefs about the condition, and coping mechanisms—should be further investigated. The list is not exhaustive and should be defined in subsequent steps, as the relevance of these factors may vary across settings. This group of patients also requires further investigations, including identifying the underlying reasons and developing strategies incorporating a multidisciplinary approach to address various aspects of the D2M situation in clinical practice.

Importantly, the criteria presented above assume the correctness of the diagnosis and patient compliance with the prescribed treatment. These are relevant aspects to consider as the first steps when dealing with D2M axSpA patients and defining TR patients. In the next step of the D2M initiative, we plan to develop recommendations on how to approach D2M and TR patients.

Our work has several limitations, which should be acknowledged. First, the approach we used was based on expert and patient opinions rather than being data-driven. While this approach is certainly less rigorous than a data-driven approach (such as that used for classification criteria), we believe that, in the absence of unified definitions and terminology at the outset of the project, this was the only feasible way to progress. A unified definition was necessary as a starting point to stimulate research and develop management recommendations for this patient group. Second, the literature review was conducted as a scoping review rather than a systematic review. However, we believe this did not compromise the work, as the goal of the review was to provide a foundation for expert consensus rather than an exhaustive synthesis of evidence. As previously mentioned, evidence synthesis would not have been possible without a unified definition and nomenclature. Third, even within the expert organisation, there was some heterogeneity of views on certain aspects of the definition, as reflected in the results of the Delphi exercises. Nonetheless, through discussion and refinement, a broad consensus was achieved, with 89% of the members endorsing the final definition.

The D2M axSpA initiative aligns well with similar efforts in other inflammatory rheumatic conditions, such as RA (termed 'difficult-to-treat') [13] and psoriatic arthritis [18,19]. In these conditions, both rheumatologists and patients often encounter challenges in achieving complete control of disease signs and symptoms, even with state-of-the-art treatments. It is anticipated that common mechanisms, such as central sensitisation, along with disease-specific factors, contribute to the development of D2M/difficult-to-treat situations. This understanding likely extends to TR disease as well, thereby stimulating research into these mechanisms across different rheumatic diseases.

In conclusion, the ASAS D2M axSpA definition allows for clear identification of patients with unmet medical needs, indicating the way forward for improved clinical management and further research.

Competing interests

DP has received research support from AbbVie, Eli Lilly, MSD, Novartis, Pfizer, consulting fees from AbbVie, Biocad,

Bristol-Myers Squibb, Eli Lilly, Janssen, Moonlake, Novartis, Pfizer, and UCB, and speaker fees from AbbVie, Canon, DKSH, Eli Lilly, Janssen, MSD, Medscape, Novartis, Peervoice, Pfizer, and UCB. DP is a member of the executive committee of ASAS. VN-C has received consultancy/speaker/research grants from AbbVie, Alphasigma, BMS, Fresenius Kabi, Galapagos, Janssen, Eli Lilly, Moonlake, MSD, Novartis, Pfizer, Roche, and UCB. VN-C is a member of the executive committee, executive secretary, and elected president of ASAS. MT: none declared. SA has received consulting fees from Argenx, Bristol-Myers Squibb, Galapagos, and Novartis. SZA has received grants from AbbVie, Eli Lilly, Jannsen, Novartis, Pfizer, and UCB, consulting fees from AbbVie, Eli Lilly, Jannsen, Novartis, Pfizer, and UCB, and honoraria from AbbVie, Eli Lilly, Jannsen, Novartis, Pfizer, and UCB. SB: none declared. FvdB: none declared. CB has received financial support from Novartis, AbbVie, Galapagos, Novartis, Pfizer, and UCB. AC has received grant support from BMS, Lilly, and Novartis, honoraria from Alfa Sigma, AbbVie, Amgen, BMS, Eli Lilly, Galapagos, Novartis, Pfizer, and UCB, and travel/meeting attendance support from AbbVie and UCB. JD: none declared. MD has received grant support from Pfizer, UCB, and AbbVie, consulting fees from Novartis and Pfizer, travel/meeting attendance support from Novartis, and participation on a Data Safety Monitoring Board or Advisory Board for Galapagos. TD has received honoraria from AbbVie, Amgen, Novartis, Koçak, and Lilly and travel/meeting attendance support from Celltrion, Lilly, AbbVie, Amgen, Novartis, and Nobel. BE-Z has received consultancy, research grants, and speaker's honoraria from AbbVie, Amgen, BMS, Eva, Hekma, Janssen, Lilly, MSD, New Bridge, Novartis, Pfizer, Roche, Sanofi-Aventis, Servier, and Sobi. WF has received speaker/consulting fees/grants from Novartis, Janssen and Janssen, GlaxoSmithKline, and Diethelm Keller Siber Hegner (DKSH). FvG has received consulting and speaker fees from AbbVie, ASAS, BMS, Galapagos, Janssen, Lilly, Novartis, Pfizer, and UCB. RG-S has received consultancy/speaker/research grants from AbbVie, BMS, Janssen, Eli Lilly, Novartis, Pfizer, Roche, UCB, GSK, Biogen, Amgen, Raffo, and Adium. MGC has received grant support and travel/meeting attendance support from Novartis. PG: none declared. LG: none declared. SG has received honoraria from AbbVie, Altamedics, Amgen, Eli Lilly, Johnson & Johnson, Krka, Medis, MSD, Novartis, Pfizer, Sandoz, Sobi, Stada, Viatris, Teva, and Zentiva and support for attending meetings/travel from AbbVie, Novartis, and Pfizer. FH: none declared. MK has received honoraria from AbbVie, Amgen, Asahi-Kasei Pharma, Ayumi Pharma, BMS, Chugai, Daiichi-Sankyo, Eisai, Eli Lilly, Gilead, Janssen, Novartis, Ono Pharma, Tanabe-Mitsubishi, and UCB Pharm. RL has received consulting fees from AbbVie, Pfizer, UCB, Eli Lilly, Novartis, Jansen Pharma, and Galapagos, honoraria from UCB and AbbVie, participated in a DSMB for a UCB trial, is ASAS and METEOR board member and director of Joint Imaging BV, and of Rheumatology Consultancy BV. YYL has received speaking fees from AbbVie, DKSH, Janssen, Novartis, and Pfizer. PMM has received honoraria from AbbVie, BMS, Celgene, Eli Lilly, Galapagos, Janssen, MSD, Novartis, Orphazyme Pfizer, Roche, and UCB. HM-O has received research grants from Janssen, Novartis, Pfizer, and UCB and speaker fees/honoraria from AbbVie, Amgen, Biogen, Eli Lilly, Janssen, Moonlake, Novartis, Pfizer, and UCB. BM has received honoraria/participated in advisory boards and/or received speaker fees from IPCA, Cipla, Torrent, RPG Lifesciences, and Novartis. AM has received speaker honoraria/participated in advisory boards, and/or received research grants from Biogen, BMS, UCB, Pfizer, Johnson & Johnson, Novartis, Amgen, Gilead, Galapagos, AbbVie, Lilly, and Pfizer. EN has received speaker honoraria/participated in advisory boards, and/or received research grants from UCB, Pfizer, Gilead, Galapagos, AbbVie, Eli Lilly, Alfasigma, Fresenius, and Pfizer. SR has received research grants and/or consultancy fees from AbbVie, Eli Lilly, Galapagos/Alfasigma, Janssen, MSD, Novartis, Pfizer, Sanofi, and UCB. MR has received consulting and/or speaker fees from AbbVie, Chugai, Boehringer-Ingelheim, Eli Lilly, Janssen, Novartis, and UCB. CGSS has received consulting and speaker fees from AbbVie, Janssen, Novartis, and Pfizer. AS has received speaking and/or consulting fees from AbbVie, Novartis, UCB, and Lilly. Support for attending conferences: Janssen. JW has received research grants, speaker fees, or advisor board from Pfizer, Abbott, AbbVie, Bristol-Myers Squibb, Eli Lilly, JNJ, UCB, MSD, GSK, Chugai, Roche, Celgene, Sanofi-Aventis, and Novartis. XB has received research support from AbbVie, Eli Lilly, and Novartis, consulting fees from AbbVie, Alphasigma, BMS, Eli Lilly, Janssen, Moonlake, Novartis, Pfizer, and UCB, and speaker fees from AbbVie, Alphasigma, BMS, Eli Lilly, Janssen, Medscape, Novartis, Peervoice, Pfizer, and UCB. President of ASAS and EULAR. DvdH has received consulting fees from AbbVie, Alfasigma, Argenx, BMS, Eli Lilly, Grey-Wolf Therapeutics, Janssen, Novartis, Pfizer, Takeda, and UCB Pharma and is director of Imaging Rheumatology bv. DvdH is the associate editor of Annals Rheumatic Diseases, an editorial board member of the Journal of Rheumatology and RMD Open, and an advisor to Assessment of Axial Spondyloarthritis International Society.

Acknowledgements

The authors thank all members of the Assessment of SpondyloArthritis International Society who contributed their valuable expertise and knowledge to the Delphi exercises, significantly influencing the refinement of the definition through engaging live discussions. HM-O is supported by the National Institute for Health Research (NIHR) Leeds Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the (UK) National Health Service, the NIHR, or the (UK) Department of Health.

Contributors

Members of the steering committee (DP, DvdH, VN-C, and XB) initiated and conducted the project, supervised the literature review, developed the surveys, and analysed their results. They also drafted the definition and the manuscript. MT conducted the literature review and supported the project's conduction and manuscript development. All authors contributed to the project's development, interpretation of survey results, drafting of the definition, and revising the manuscript for important intellectual content. All authors have approved the final version of the manuscript. DP acts as the guarantor author of this manuscript.

Funding

The project has been conducted under the auspices and with the support of the Assessment of SpondyloArthritis International Society (ASAS). The work of MT was supported by an ASAS research grant.

Patient consent for publication

This project actively involved patients with axial spondyloarthritis (axSpA) throughout the study design, consensus-building process, and the development of the final definition of difficult-to-manage axSpA. Two patient representatives were included as coauthors and contributed significantly to all stages of the project.

Ethics approval

No ethics committee approval was required for this study.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.035.

Orcid

Denis Poddubnyy: http://orcid.org/0000-0002-4537-6015
Victoria Navarro-Compán: http://orcid.org/0000-0002-4527-852X
Murat Torgutalp: http://orcid.org/0000-0003-4600-9484
Suzanne Arends: http://orcid.org/0000-0002-4422-7640
Sibel Zehra Aydin: http://orcid.org/0000-0001-8792-4449
Simone Battista: http://orcid.org/0000-0002-7471-1951
Filip van den Bosch: http://orcid.org/0000-0002-3561-5932
Christine Bundy: http://orcid.org/0000-0002-5981-3984
Alberto Cauli: http://orcid.org/0000-0002-7955-6786
Maxime Dougados: http://orcid.org/0000-0003-3009-6229
Tuncay Duruöz: http://orcid.org/0000-0003-3584-2788
Bassel El-Zorkany: http://orcid.org/0000-0003-2704-9712
Warren Fong: http://orcid.org/0000-0003-1891-1892
Floris van Gaalen: http://orcid.org/0000-0001-8448-7407
Rodrigo Garcia-Salinas: http://orcid.org/0000-0002-5928-1092
Marco Garrido Cumbrera: http://orcid.org/0000-0001-9727-1189
Pál Géher: http://orcid.org/0000-0002-4543-9629
Lianne Gensler: http://orcid.org/0000-0001-6314-5336
Simeon Grazio: http://orcid.org/0000-0003-3407-0317
Feng Huang: http://orcid.org/0000-0002-2319-873X
Mitsumasa Kishimoto: http://orcid.org/0000-0002-4007-1589
Robert Landewé: http://orcid.org/0000-0002-0577-6620
Ying Ying Leung: http://orcid.org/0000-0001-8492-6342
Pedro M Machado: http://orcid.org/0000-0002-8411-7972
Helena Marzo-Ortega: http://orcid.org/0000-0002-9683-3407
Bhowmik Meghnathi: http://orcid.org/0000-0001-7337-5194
Anna Molto: http://orcid.org/0000-0003-2246-1986
Elena Nikiphorou: http://orcid.org/0000-0001-6847-3726
Sofia Ramiro: http://orcid.org/0000-0002-8899-9087
Martin Rudwaleit: http://orcid.org/0000-0001-5445-548X
Carla G S Saad: http://orcid.org/0000-0001-9720-0719
Alexandre Sepriano: http://orcid.org/0000-0003-1954-0229
James Wei: http://orcid.org/0000-0002-1235-0679
Xenofon Baraliakos: http://orcid.org/0000-0002-9475-9362
Désirée van der Heijde: http://orcid.org/0000-0002-5781-158X

REFERENCES

[1] Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017;390(10089):73–84.
[2] Navarro-Compán V, Sepriano A, El-Zorkany B, van der Heijde D. Axial spondyloarthritis. Ann Rheum Dis 2021;80(12):1511–21.
[3] van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. A proposal for modification of the New York criteria. Arthritis Rheum 1984;27(4):361–8.
[4] van der Heijde D, Molto A, Ramiro S, Braun J, Dougados M, van Gaalen FA, et al. Goodbye to the term 'ankylosing spondylitis', hello 'axial spondyloarthritis': time to embrace the ASAS-defined nomenclature. Ann Rheum Dis 2024;83(5):547–9.
[5] Ramiro S, Nikiphorou E, Sepriano A, Ortolan A, Webers C, Baraliakos X, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. Ann Rheum Dis 2023;82(1):19–34.
[6] Smolen JS, Braun J, Dougados M, Emery P, Fitzgerald O, Helliwell P, et al. Treating spondyloarthritis, including ankylosing spondylitis and psoriatic arthritis, to target: recommendations of an international task force. Ann Rheum Dis 2014;73(1):6–16.
[7] Machado PM, Landewé R, Heijde DV, Assessment of SpondyloArthritis international Society (ASAS). Ankylosing Spondylitis Disease Activity Score (ASDAS): 2018 update of the nomenclature for disease activity states. Ann Rheum Dis. 2018;77(10):1539–40.
[8] Sieper J, Rudwaleit M, Baraliakos X, Brandt J, Braun J, Burgos-Vargas R, et al. The Assessment of SpondyloArthritis International Society (ASAS) handbook: a guide to assess spondyloarthritis. Ann Rheum Dis 2009;68(suppl 2) ii1-44.
[9] Webers C, Ortolan A, Sepriano A, Falzon L, Baraliakos X, Landewé RBM, et al. Efficacy and safety of biological DMARDs: a systematic literature review informing the 2022 update of the ASAS-EULAR recommendations for the management of axial spondyloarthritis. Ann Rheum Dis 2023;82(1):130–41.
[10] Al Mohamad F, Rios Rodriguez V, Haibel H, Protopopov M, Rademacher J, Sieper J, et al. Association of nociplastic and neuropathic pain components with the presence of residual symptoms in patients with axial spondyloarthritis receiving biological disease-modifying antirheumatic drugs. RMD Open 2024;10(1):e004009.
[11] Kieskamp SC, Paap D, Carbo MJG, Wink F, Bos R, Bootsma H, et al. Central sensitization, illness perception and obesity should be considered when interpreting disease activity in axial spondyloarthritis. Rheumatology (Oxford) 2021;60(10):4476–85.
[12] Kieskamp SC, Paap D, Carbo MJG, Wink F, Bos R, Bootsma H, et al. Central sensitization has major impact on quality of life in patients with axial spondyloarthritis. Semin Arthritis Rheum 2022;52:151933.
[13] Nagy G, Roodenrijs NMT, Welsing PM, Kedves M, Hamar A, van der Goes MC, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2021;80(1):31–5.
[14] Nagy G, Roodenrijs NMT, Welsing PMJ, Kedves M, Hamar A, van der Goes MC, et al. EULAR points to consider for the management of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2022;81(1):20–33.
[15] Baraliakos X, Listing J, von der Recke A, Braun J. The natural course of radiographic progression in ankylosing spondylitis—evidence for major individual variations in a large proportion of patients. J Rheumatol 2009;36(5):997–1002.
[16] Torgutalp M, Rios Rodriguez V, Dilbaryan A, Proft F, Protopopov M, Verba M, et al. Treatment with tumour necrosis factor inhibitors is associated with a time-shifted retardation of radiographic spinal progression in patients with axial spondyloarthritis. Ann Rheum Dis 2022;81(9):1252–9.
[17] Poddubnyy D, Sieper J. Mechanism of new bone formation in axial spondyloarthritis. Curr Rheumatol Rep 2017;19(9):55.
[18] Marzo-Ortega H, Harrison SR, Nagy G, Machado PM, McGonagle DG, Aydin SZ, et al. Time to address the challenge of difficult to treat psoriatic arthritis: results from an international survey. Ann Rheum Dis 2024;83(3):403–4.
[19] Ribeiro AL, Singla S, Chandran V, Chronis N, Liao W, Lindsay C, et al. Deciphering difficult-to-treat psoriatic arthritis (D2T-PsA): a GRAPPA perspective from an international survey of healthcare professionals. Rheumatol Adv Pract 2024;8(3):rkae074.

Ann Rheum Dis 84 (2025) 547–553

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Rheumatoid arthritis

Predictors of arthritis development in individuals at risk of rheumatoid arthritis: a 5-year follow-up study from a large cohort

Giulia Frazzei[1,2,*], Robert B.M. Landewé[3], Carlijn Wagenaar[3,4], Lotte A. van de Stadt[4], Dirkjan van Schaardenburg[3,4], Sander W. Tas[1,2], Ronald F. van Vollenhoven[1]

[1] *Department of Rheumatology and Clinical Immunology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands*
[2] *Department of Experimental Immunology, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands*
[3] *Amsterdam Rheumatology and Immunology Center, Amsterdam, Netherlands*
[4] *Amsterdam Rheumatology and Immunology Center, Reade, Amsterdam, Netherlands*

ARTICLE INFO

Article history:
Received 17 April 2024
Received in revised form 8 January 2025
Accepted 8 January 2025

ABSTRACT

Objectives: The aim of this study was to predict rheumatoid arthritis (RA) development in a cohort of at-risk individuals with arthralgia positive for anti-citrullinated protein antibodies (ACPA) and/or rheumatoid factor (IgM-RF), followed for up to 5 years.

Methods: In total, 617 seropositive arthralgia individuals were included in the study. The ability of clinically and biologically relevant baseline characteristics to predict RA development was assessed using Cox proportional hazard regression analysis.

Results: Thirty-eight percent of study population was IgM-RF-positive, 31% was ACPA-positive, and 30% was positive for both ACPA and IgM-RF. Mean (SD) time till arthritis was 19.6 (19.0) months in 33.7% of participants; mean (SD) follow-up time of individuals who did not develop arthritis was 47.3 (24.5) months. We found that first-degree relatives of RA (hazard ratio [HR] = 1.50), individuals who had intermittent symptoms (HR = 1.64), symptoms for less than 12 months at inclusion (HR for symptom duration >12 months = 0.71), morning stiffness ≥1 hour (HR = 1.63), or reported joint swelling (HR = 1.51) independently had higher risk to develop arthritis. Moreover, individuals with high ACPA titres (HR = 4.65) or double positivity for ACPA and IgM-RF (HR = 6.83) had the highest risk of developing RA, as compared to those with only IgM-RF or low ACPA titres. The risk of developing arthritis was 58.2% when at least 3 variables were present.

Conclusions: Baseline characteristics can be used to predict future RA development in seropositive arthralgia individuals. These results will aid in the identification of individuals at highest risk of developing RA, who could potentially benefit from additional follow-ups in clinical practice and recruitment in preventive trials.

*Correspondence to Giulia Frazzei.
E-mail address: g.frazzei@amsterdamumc.nl (G. Frazzei).
Handling editor Josef S Smolen.

https://doi.org/10.1016/j.ard.2025.01.042

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Rheumatoid arthritis (RA) is preceded by a pre-clinical stage in which individuals experience nonspecific signs and symptoms.
- The presence of antibodies against citrullinated proteins (anti-citrullinated protein antibodies [ACPA]) and rheumatoid factor (IgM-RF) can be detected years before the first symptom manifestations and has been found to be associated with future arthritis development in several at-risk cohorts.
- One of the challenges of preventive trials in RA-risk individuals is the identification of individuals who would benefit the most from intervention.

WHAT THIS STUDY ADDS

- Being first-degree relative of patients with RA, having symptoms for less than 1 year at time of inclusion, having intermittent symptoms, reporting swollen joints and morning stiffness for longer than 1 hour, and being highly positive for ACPA or double positive for both ACPA and IgM-RF predicted an higher risk of developing arthritis.
- Age, gender, or smoking status were not predictors of developing arthritis in our cohort.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- This study confirms the robustness of 4 variables in predicting the risk of developing RA. Those variables could be used in patient recruitment strategies in future preventive trials.

INTRODUCTION

Rheumatoid arthritis (RA) is an autoimmune disease characterised by joint damage and systemic features, which has a detrimental impact on patients' life if left untreated [1,2]. Its onset is preceded in many patients by an asymptomatic phase, in which genetic and environmental risk factors contribute to recognition of autoantigens, induction of autoreactive humoral immune responses, and low-grade inflammation. One of the most relevant genetic risk factors is a specific allele, called shared epitope (SE) in the HLA-DRB1 locus, with high affinity for citrullinated proteins [3–5]. In this pre-clinical stage, individuals could manifest nonspecific signs and symptoms, such as join pain (arthralgia) and morning stiffness [6–8]. In prospectively collected biorepositories, the presence of circulating autoantibodies, such as anti-citrullinated protein antibodies (ACPA) and rheumatoid factor (IgM-RF), was detected years before disease onset in some patients [9]. Moreover, presence of such antibodies is considered a risk factor for development of RA in individuals with arthralgia. However, not all seropositive arthralgia individuals will eventually develop RA [10,11]. Several studies have investigated the performance of biological markers, clinical manifestations, and imaging modalities in predicting RA development, including genetic and environmental risk factors (ie, SE, *PTPN22*, *CTLA4*, smoking, and stress); circulating autoantibodies (ie, ACPA and IgM-RF); alterations in the immune cell frequency (ie, anomalies in T cell subsets); physical manifestations (ie, arthralgia, fatigue, and morning stiffness); and subclinical inflammation as detected by imaging techniques (ie, ultrasound [US] and/or magnetic resonance imaging [MRI]), with conflicting results [3,5,10,12–19]. As of now, consensus on how to predict RA development in at-risk individuals is lacking.

Effective preventive strategies would benefit individuals and reduce the socioeconomic impact of RA. These interventions may be aimed at specific stages of disease pathogenesis: that is, either complete prevention with a focus on eliminating disease risk before autoimmunity onset or prevention of clinical disease development, in which at-risk individuals would have to be treated before manifestation of clinical signs or symptoms. In the past years, numerous studies have investigated the effectiveness of preventing RA in its at-risk stage, with varying results [20]. Profound $CD20^+$ B cell depletion through a single dose of rituximab in 81 seropositive arthralgia individuals led to 55% lower background risk of RA development at 12 months after treatment, although it did not prevent the development of RA [21]. While 1-year treatment with methotrexate also did not significantly prevent RA in 236 arthralgia individuals with MRI-detected subclinical inflammation, there was a significant improvement in patients' perception of disease burden over time [22]. The treatment of ACPA-positive arthralgia individuals with abatacept for 6 to 12 months led to a decrease in the rate of RA development at the end of treatment, with sustained, although less marked, differences at 12 months after the treatment, that is, less pain and improved quality of life [23,24].

Considering that individuals with clinically suspected arthralgia experience already a serious disease burden, improvement in quality of life in this stage is important and should perhaps be considered just as relevant as prevention of RA [25]. A key challenge in this context is the identification of those RA-risk individuals who would benefit the most from preventive treatment. Since this is as yet not possible, many prevention studies suffer from recruitment of a heterogeneous study population, which may yield results that are unclear or difficult to interpret. Recently, a European League Against Rheumatism (EULAR) taskforce has published considerations on conducting clinical trials and observational studies in RA-risk individuals, in which they address the importance of defining study populations correctly [26]. A systematic literature review on risk factors for arthritis development was recently published by members of this taskforce to aid in the guideline decision making [27].

RA-risk individuals could either be asymptomatic individuals, who do not have any clinical manifestations and are often first-degree relatives (FDRs) of patients with RA, or at-risk individuals with nonspecific signs and symptoms without clinical arthritis. EULAR recommendations for conducting clinical studies in this population define seropositive RA-risk individuals with symptoms without clinical RA as individuals with musculoskeletal symptoms, arthralgia, or clinically suspect arthralgia (CSA) according to their symptom manifestations [26]. In 2013, van de Stadt *et al* published a prediction model for RA development in a RA-risk cohort in which seropositive individuals with arthralgia were stratified into 3 categories (low, intermediate, and high) based on a given score [17]. This prediction model was developed with data of 374 subjects recruited at rheumatology outpatient clinics in Amsterdam, Netherlands, and we have now additional data from 243 individuals. The aim of this study was to update the risk of arthritis development in the expanded cohort and the performance of the previously reported risk factors 10 years after publication.

METHODS

Study population

Individuals with arthralgia (defined as pain in any extremity joint, as reported by study subjects) and who tested positive for ACPA and/or IgM-RF were recruited between 2004 and 2019 at the rheumatology outpatient clinics at Reade in Amsterdam,

Netherlands. Study participants were required to have absence of arthritis at physical examination, not have received glucocorticoids in the past 3 months, and had to be naïve of disease modifying anti-rheumatic drugs (DMARDs) [17]. Within study duration, participants were allowed treatment with nonsteroidal anti-inflammatory drugs; initiation of glucocorticoid or DMARD treatment after study initiation was not a reason for exclusion from the study.

Patient and public involvement

Patients or the public were not involved in the design, conduct, or reporting of our research. We will involve patient research partners in the dissemination of current work to the public, as well as to patient groups.

Study schedule, data collection, and study outcome

Study subjects were seen at 6 and 12 months for the first year of the study, and then annually for the duration of the study, up to 5 years follow-up. At each study visit, physical examination and routine blood tests were performed, and data were collected on smoking habits and alcohol consumption, joint complaints, tender joints, pain as reported through visual analogue scale (VAS), morning stiffness, frequency and duration of symptoms. In addition, at baseline data were recorded on medical history and family history of RA, and genetic tests were performed. Participants were excluded from the study if they presented with clinical arthritis at baseline. Arthritis was defined as at least 1 swollen joint, as independently evaluated by 2 rheumatologists. Imaging, such as US or MRI, was not used to establish the presence or absence of arthritis. When arthritis development was suspected, an additional visit was performed. Arthritis development was considered present only when 2 physicians independently confirmed the same swollen joint(s).

Laboratory analysis

Baseline levels of ACPA and IgM-RF were determined by second-generation aCCP ELISA (Axis Shield) and inhouse ELISA, as previously described [9]. The cut-off level for ACPA positivity was set according to manufacturer's instruction at 5 arbitrary units/mL (AU/mL). For IgM-RF positivity, cut-off was set at 30 IU/mL based on analysis of receiver operating characteristic (ROC) curves from 2 cohorts, as reported by Nielen *et al* [9]. C-reactive protein (CRP) was analysed through a highly sensitive latex-enhanced assay on a Hitachi 911 analyser (Roche Diagnostics; assay range 0.15-200 mg/L); cut-off was set according to manufacturer's instruction at 10 mg/L. Genetic analyses to infer SE carrier status were performed as previously described by van de Stadt *et al* [17].

Statistical analysis

Statistical analysis were performed with IBM SPSS Statistics 28. Variables were selected based on their clinical and biological applicability, as reported in the literature. To estimate the probability of arthritis development based on selected variables (age, gender, the presence of reported swollen joint(s), smoking, duration of symptoms, type of symptoms, VAS pain, CRP positivity, and antibody status), multivariable Cox proportional hazard regression analysis was performed. Although clinically and biologically possibly relevant, positivity for SE was excluded from Cox proportional hazard analysis due to a too high proportion of missing values (nearly 40% of total). When relevant, continuous variables were categorised based on clinically applicable cut-off or percentile. Antibody status (IgM-RF and ACPA positivity) was analysed as a categorical variable with 4 categories: IgM-RF positive, ACPA low, ACPA high, and positive for both ACPA and IgM-RF. ACPA low was defined as aCCP levels higher than upper limit of normal (ULN) but less than $<3 \times$ ULN, ACPA high was defined as aCCP levels $\geq 3 \times$ ULN. IgM-RF positivity was not further stratified in low and high, since no significant difference was found in the distribution of IgM-RF titres.

RESULTS

Population overview and arthritis development

Out of 659 seropositive arthralgia individuals included in Reade between 2004 and 2019, 617 subjects were followed up for arthritis development and analysed in this study; 41 patients were lost to follow-up over time. The majority of the study population were women (459 individuals, 74.4%), with a mean age of 49.7 (±11.6) years. Two hundred and thirty-eight (38.8%) subjects were IgM-RF positive, 189 (30.8%) subjects were ACPA positive, and 187 (30.4%) subjects were positive for both IgM-RF and ACPA; data on 3 participants were missing at baseline. When we stratified for the levels of ACPA, 107 (17.4%) subjects had low antibody levels (ACPA $<3 \times$ ULN), while 269 (43.8%) had high antibody levels (ACPA $\geq 3 \times$ ULN). We found a small but significant association between smoking and ACPA positivity (odd ratio [OR] = 1.46; P value = .047). Moreover, of 377 individuals whose genetic data were available, 205 (54.4%) were positive for SE. A summary of study baseline characteristics is reported in Table 1; baseline characteristics, stratified by year of inclusion, are reported in Supplementary Table S1.

Two hundred and eight subjects (33.7%) developed arthritis, with a mean (SD) time until arthritis of 19.6 (19.0; median [IQR]: 13 [6-27]) months, and median (IQR) SJC44, as evaluated by rheumatologist, of 3 (2-6) at time of arthritis. Mean (SD) follow-up time for individuals who did not develop arthritis was 47.3 (24.5; median [IQR]: 58 [25-61]) months. Of those individuals who developed arthritis, 174 (83.7%) subjects met the classification criteria for RA and 11 (5.3%) the classification criteria for undifferentiated arthritis (UA) according to the 2010 American College of Rheumatology (ACR)/EULAR criteria [28]. Baseline characteristics, stratified based on arthritis development, are reported in Table 2.

Prediction of arthritis development

To investigate which factors predicted arthritis development in our study population, we performed a multivariable Cox proportional hazard regression analysis including clinically relevant variables, as well as risk factors known from the literature, such as age, gender, smoking, reported swollen joint count, duration and type of symptoms, VAS pain and CRP positivity, and antibody status (Table 3). Survival curve of arthritis progression for overall subjects and individuals stratified based on year of inclusion is shown in Supplementary Figure S1. The baseline risk of developing arthritis at 1, 3, and 5 years have been reported in Supplementary Table S2.

Out of the analysed parameters, being an FDR of patients with RA, having symptoms for less than 1 year at time of inclusion, having intermittent symptoms, reported swollen joints, morning stiffness ≥1 hour, and subjects' antibody status were found to be significantly associated with the risk of developing

Table 1
Baseline characteristics of study cohort

Variables	Values	No. missing values
Age in years, mean (±SD)	49.7 (±11.6)	
Gender female, n (%)	459 (74.4%)	—
FDR with RA, n (%)	135 (21.9%)	—
Current smoker, n (%)	167 (27.2%)	2
Alcohol use, n (%)	370 (61.6%)	17
Symptom duration >12 months, n (%)	403 (66.4%)	10
Intermittent symptoms, n (%)	193 (32.1%)	16
VAS pain ≥50, n (%)	209 (32.1%)	—
Morning stiffness ≥1h, n (%)	145 (23.5%)	1
TJC53, n (%)		—
No tender joint	311 (50.4%)	
1-5 tender joints	215 (34.8%)	
6-10 tender joints	62 (10.0%)	
>11 tender joints	29 (4.7%)	
TJC53		—
Mean (SD)	2.5 (4.8)	
Median (IQR)	0 (0-3)	
Reported SJC44, n (%)		—
no swollen joint	401 (65.0%)	
1-5 swollen joints	159 (25.8%)	
6-10 swollen joints	39 (6.3%)	
>11 swollen joints	18 (2.9%)	
Reported SJC44		—
Mean (SD)	1.7 (4.0)	
Median (IQR)	0 (0-2)	
CRP positive, n (%)	57 (9.3%)	—
Antibody status, n (%)		3
IgM-RF positive	238 (38.8%)	
aCCP low positive	80 (13.0%)	
aCCP high positive	109 (17.8%)	
aCCP and IgM-RF positive	187 (30.4%)	
SE positive, n (%)	205 (54.4%)	240
Arthritis development (yes), n (%)	208 (33.7%)	—

Baseline characteristics are reported as number (%), mean (±SD), or median (IQR). SJC44 refers to swollen joint count as reported by study participants.
aCCP high, ACPA ≥3 × upper limit of normal; aCCP low, ACPA <3 × upper limit of normal; aCCP, anti-cyclic citrullinated peptide antibody; ACPA, anti-citrullinated protein antibodies; CRP, C-reactive protein; FDR, first-degree relatives; IgM-RF, rheumatoid factor; SE, shared epitope; SJC44, swollen joint count 44; TJC53, tender joint count 53; VAS, visual analogue scale.

Table 2
Baseline characteristics of individuals who developed arthritis vs individuals who did not develop arthritis

Variables	Arthritis yes (n = 208)	Arthritis no (n = 409)
Age in years, mean (±SD)	49.0 (±11.5)	50.1 (±11.6)
Gender female, n (%)	153 (73.6%)	306 (74.8%)
FDR with RA, n (%)	64 (30.8%)	71 (17.4%)
Current smoker, n (%)	61 (29.5%)	106 (26.0%)
Alcohol use, n (%)	114 (57.0%)	256 (64.0%)
Symptom duration >12 months, n (%)	125 (61.0%)	278 (69.2%)
Intermittent symptoms, n (%)	94 (46.1%)	99 (24.9%)
VAS pain ≥50, n (%)	80 (38.5%)	129 (31.5%)
Morning stiffness ≥1h, n (%)	53 (25.6%)	92 (22.5%)
TJC53, n (%)		
No tender joint	108 (51.9%)	203 (49.6%)
1-5 tender joints	71 (34.1%)	144 (35.2%)
6-10 tender joints	22 (10.6%)	40 (9.8%)
>11 tender joints	7 (3.4%)	22 (5.4%)
TJC53		
Mean (SD)	2.3 (3.9)	2.6 (5.2)
Median (IQR)	0 (0-3)	1 (0-3)
Reported SJC44, n (%)		
No swollen joint	106 (51.0%)	295 (72.1%)
1-5 swollen joints	74 (35.6%)	85 (20.8%)
6-10 swollen joints	19 (9.1%)	20 (4.9%)
>11 swollen joints	9 (4.3%)	9 (2.2%)
Reported SJC44		
Mean (SD)	2.4 (4.6)	1.3 (3.6)
Median (IQR)	0 (0-3)	0 (0-1)
CRP positive, n (%)	25 (12.3%)	32 (7.8%)
Antibody status, n (%)		
IgM-RF positive	26 (12.6%)	212 (52.1%)
aCCP low positive	22 (10.6%)	58 (14.3%)
aCCP high positive	48 (23.2%)	61 (15.0%)
aCCP and IgM-RF positive	111 (53.6%)	77 (18.7%)
SE positive, n (%)	92 (60.9%)	113 (50.0%)

SJC44 refers to swollen joint count as reported by study participants.
aCCP high, ACPA ≥3 × upper limit of normal; aCCP low, ACPA <3 × upper limit of normal; aCCP, anti-cyclic citrullinated peptide antibody; ACPA, anti-citrullinated protein antibodies; CRP, C-reactive protein; FDR, first-degree relatives; IgM-RF, rheumatoid factor; SE, shared epitope; SJC44, swollen joint count 44; TJC53, tender joint count 53; VAS, visual analogue scale.

Table 3
Multivariable Cox proportional hazard regression analysis

	HR	95% CI	*P value*
Age in years	1.01	1.00, 1.02	.23
Gender female	1.06	0.75, 1.50	.73
FDR with RA	1.50	1.09, 2.07	.013
Current smoker	0.85	0.61, 1.19	.35
Alcohol consumption	0.77	0.56, 1.05	.10
Symptom duration >12 months	0.71	0.52, 0.95	.023
Intermittent symptoms	1.64	1.19, 2.25	.002
VAS pain ≥50	1.33	0.98, 1.80	.07
Reported SJC44	1.51	1.12, 2.02	.007
Morning stiffness ≥1 h	1.63	1.15, 2.30	.006
CRP positive	1.26	0.80, 1.99	.32
Antibody status			<.001
IgM-RF positive	Reference		
aCCP low positive	2.21	1.20, 4.05	.01
aCCP high positive	4.65	2.78, 7.79	<.001
aCCP and IgM-RF positive	6.83	4.27, 10.93	<.001

HR, CI, and *P* values are reported for each variable analysed. All variables, with the exception of age, were analysed as categorical variables.
aCCP high, ACPA ≥3 × upper limit of normal; aCCP low, ACPA <3 × upper limit of normal; aCCP, anticyclic citrullinated peptide antibody; ACPA, anti-citrullinated protein antibodies; CI, confidence interval; CRP, C-reactive protein; FDR, first-degree relatives; HR, hazard ratio; IgM-RF, rheumatoid factor; SJC44, swollen joint count 44; TJC53, tender joint count 53; VAS, visual analogue scale.

arthritis. In particular, having high ACPA titres or being double positive for ACPA and IgM-RF were predictor of increased risk. The absolute risk of developing RA when at least 4 of the variables are present at baseline is 67.6% (46 individuals with RA out of 68 in total) (Table 4).

DISCUSSION

Not all seropositive arthralgia individuals will develop RA; however, at present, no formally endorsed guidelines are available for the prediction of arthritis in this population. While the presence of autoantibodies is a risk factor for developing RA, this biomarker on its own is not enough to differentiate individuals with high risk of disease development from those who will never develop RA. In this study, we investigated a cohort of seropositive arthralgia individuals and identified 6 variables associated with high risk of developing arthritis. Our study population was primarily female (74.4%), which might indicate a role of gender in the immune activation phase, before sign and symptom manifestations.

Considering the strong impact of arthritis on patients' quality of life and the impairment of physical abilities, our primary

Table 4
Absolute risk of arthritis development

Variables	TP	FP	TN	FN	Sensitivity	Specificity	PPV	NPV
≥3	117	84	306	83	58.5	78.5	58.2	78.7
≥4	46	22	368	154	23.0	94.4	67.6	70.5
≥5	10	0	390	190	5.0	100	100	67.2
6	2	0	390	198	1.0	100	100	66.3

Number of individuals who are positive for at least 3, 4, 5, or all predictive variables at baseline, as well as PPV and negative predictive value (NPV) of arthritis development for selected variables (FDR with RA, symptoms <1 year, intermittent symptoms, having reported swollen joint(s), morning stiffness ≥1 hour, high ACPA titres or double positive for ACPA and IgM-RF).
ACPA, anti-citrullinated protein antibodies; FDR, first-degree relatives; FN, false negative; FP, false positive; high ACPA, ACPA ≥3 × upper limit of normal; NPV, negative predictive value; PPV, positive predictive value; IgM-RF, rheumatoid factor; TN, true negative; TP, true positive.

outcome included both UA and RA, defined according to the ACR/EULAR 2010 definition. Limited data availability on radiographic progression prohibited detailed analysis on erosive RA, which was thus not considered as an outcome for our analysis. Moreover, since bone erosion is almost never present in very early at-risk individuals, this variable was not an appropriate candidate for arthritis prediction in this cohort. Although no difference between those individuals who did and did not progress to arthritis was found in the physician-assessed number of tender joints at baseline, those individuals who developed arthritis later on reported joint swelling more often themselves, as well as more swollen joints at baseline, compared with those who did not develop arthritis. This suggests that perceived disease burden may precede clinically detectable arthritis for years. In individuals with numerous tender joints, the similarly high levels in non-progressors vs progressors may be indicative of a pain syndrome mimicking pre-arthritis in the non-progressors. For future studies, it might be of interest to also include imaging techniques to monitor disease progression over time.

Our analysis identified several predictors of arthritis development in individuals with the highest risk of developing RA. However, while these predictors could be useful to discriminate high and low risk subjects, they might not be so effective in individuals with intermediate risk. RA-risk individuals positive for some, but not all, of the identified predictor factors could also benefit from additional monitoring and tests, such as US and MRI [29]. Our data suggest that a strategy of screening FDRs of patients with RA could readily identify a subset—based mostly on serology—with a >80% risk of developing RA. These individuals would in most instances be motivated to participate in treatment protocols aimed at preventing the disease.

We reported data from a longitudinal cohort of seropositive arthralgia individuals. Part of this cohort has been analysed by van de Stadt *et al* and a prediction model was derived from 374 individuals [17]. In this study, we report additional follow-up time for included individuals, as well as an increase in the cohort population. We identified 6 variables significantly associated with higher risk of arthritis onset in our cohort: FDR of patients with RA, the presence of intermittent symptoms, symptom duration for less than 12 months at time of inclusion, morning stiffness ≥1 hour, the presence of swollen joints as reported by study participants, and either high ACPA titres or double positivity for ACPA and IgM-RF. The incidence of arthritis seemed to decrease by calendar year. Speculatively, this may be due to reduced smoking over time or changes in life style.

Antibody status, intermittent symptoms, and morning stiffness were consistently contributory to explaining arthritis risk, a risk that persisted after internal validation by randomly dividing the cohort into halves. However, FDR of patients with RA, the presence of intermittent symptoms, and reporting of swollen joints by patients themselves at baseline were only weakly statistically contributory and in only 1 of both halves. This suggests that these variables have only marginal predictive value in comparison with those with high predictive value, such as antibody status, something that also came out of our primary analysis. Of note, since we only included variables that were statistically significant in our analysis, our model may have missed variables with relatively low predictive value. While it might not be feasible in prevention trials to limit the inclusion to those individuals in whom all risk factors are present, the presence of at least 4 risk factors was highly predictive of arthritis development (67% in our cohort), and such algorithm could be used for the selection of study participants in whom the risk of RA is high enough to justify preventive DMARD treatment.

Our results are consistent with the reported association between arthritis risk and antibody status, and in particular with the notion that high ACPA titres and/or the presence of both ACPA and IgM-RF are associated with increased risk of arthritis onset [10,11,27,30]. While many of the studies analysing antibody status alone included only asymptomatic individuals, our cohort was composed of subjects with arthralgia; this suggests that the presence of ACPA is predictive of arthritis early in the autoimmune process and maintains a role even after the first (nonspecific) symptom manifestations. In contrast, the presence of IgM-RF alone was not indicative of an increased risk of arthritis in our cohort, which is consistent with the findings from Bos *et al* [10]. Since our cohort only included seropositive participants, it was not possible to evaluate the risk of arthritis in ACPA-positive and/or IgM-RF−positive subjects compared with those presenting with seronegative arthralgia, a fact that limits our findings to seropositive arthralgia individuals.

Our data also confirmed an association between being FDR of patients with RA, reporting intermittent symptoms, swollen joints at inclusion, or the presence of morning stiffness for longer than 1 hour, and the risk of arthritis onset [17,30,31]. The latter association was not found in the Leiden cohort, in which morning stiffness was not predictive of arthritis development [18]. This difference might be due to the differences in the population studied, since only subjects who reported experiencing arthralgia for less than 1 year were included in the Leiden cohort, while symptom duration was not an inclusion criteria in our population. This might also explain why CRP positivity, which was a risk factor in the Leiden cohort, was not associated with arthritis development in our cohort. Considering that the presence of symptoms for less than 1 year was predictive of arthritis onset in our cohort, this might indicate a link between detection of subclinical inflammation via CRP measurement and the manifestation of nonspecific symptoms, with limited contribution of a single CRP measurement to arthritis prediction in subjects with symptoms for more than 1 year.

In the prediction model published by van de Stadt *et al*, alcohol consumption was reported to be protective for RA development in RA-risk individuals [17]. However, in our analysis, alcohol consumption was not significantly associated with arthritis development. This difference in results could be due to a less strong effect of this variable in our cohort. Moreover, our data set lacks data on the exact amount of alcohol consumption, which may have limited our analysis. Smoking was also not found to significantly predict arthritis development. Considering that smoking is associated with citrullination of peptides in the lung, which could lead to autoantibody formation [32], it might

be that smoking does not relevantly contribute to the prediction of RA development in individuals who already have autoantibodies. Nevertheless, smoking should still be considered an important risk factor in FDRs of patients with RA who have yet to develop autoantibodies.

Similarly, SE positivity is associated with higher risk of RA development in relation to the increased risk of ACPA production. Our population consisted of seropositive arthralgia individuals, who have already developed an autoreactive humoral immune response and clinical symptoms. It is likely that SE positivity would have been associated with an even higher risk of disease development in individuals with high ACPA titres. However, when we limited our analysis to those individuals for whom SE was assessed, SE was not contributory to disease risk, not even with a trend towards developing arthritis more often. Considering the substantial number of missing values for SE positivity in our cohort and its collinearity with ACPA, we refrained from performing multiple imputation of SE, which is why this variable was omitted from further analysis.

In summary, our analysis identified 4 robust variables easily assessed in the clinic that are significantly associated with RA development. The identification of predictors for RA development can aid clinicians and scientists in the selection of those individuals who would benefit the most from preventive strategies, as well as those who might require additional testing and follow-up. The stratification of arthralgia individuals based on their personal risk might aid in the selection of subjects eligible for preventive clinical trials. This will ultimately benefit not only the individuals but also society.

Competing interests

GF, RBML, CW, LAvdS, and DvS report no conflict of interest. RFvV reports institutional grants from AstraZeneca, BMS, Galapagos, MSD, Novartis, Pfizer, Roche, Sanofi, and UCB, as well as institutional fees from AbbVie, AstraZeneca, Biogen, BSM, Galapagos, GSK, Janssen, Pfizer, RemeGen, and UCB, all outside the submitted work. SWT reports institutional grants from GSK, Lilly, Celgene, Pfizer, Roche, AstraZeneca, Galapagos, Citryll, and Galvani bioelectronics, as well as institutional fees from NovoNordisk, Pfizer, AbbVie, and UCB, all outside the submitted work.

Acknowledgements

We gratefully thank all study participants for their effort. This work has been presented at EULAR 2024, under the abstract number POS0664.

Contributors

RFvV contributed to conceptualisation of the study, data interpretation, and writing of the report. SWT, LAvdS, and DvS contributed to data interpretation and writing of the report. RBML contributed to data analysis and data interpretation. GF contributed to data collection, data analysis, data interpretation, and writing of the report. CW contributed to the data collection. All authors were involved in the drafting or revising of the intellectual content of the paper, and all authors approved the final version to be published. RFvV is the guarantor.

Funding

This work was supported by the Dutch Arthritis Association [Grant number 0801034] and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant [Agreement no. 847551] (to GF).

Patient consent for publication

All participants gave informed consent before taking part in the study.

Ethics approval

This study was approved by the Ethics committee of the Slotervaart Hospital and Reade, Amsterdam, Netherlands.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.042.

REFERENCES

[1] Scherer HU, Häupl T, Burmester GR. The etiology of rheumatoid arthritis. J Autoimmun 2020;110:102400.
[2] van de Sande MG, de Hair MJ, van der Leij C, Klarenbeek PL, Bos WH, Smith MD, et al. Different stages of rheumatoid arthritis: features of the synovium in the preclinical phase. Ann Rheum Dis 2011;70(5):772–7.
[3] Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, et al. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. Am J Hum Genet 2002;71(3):585–94.
[4] Lundström E, Källberg H, Alfredsson L, Klareskog L, Padyukov L. Gene-environment interaction between the DRB1 shared epitope and smoking in the risk of anti-citrullinated protein antibody-positive rheumatoid arthritis: all alleles are important. Arthritis Rheum 2009;60(6):1597–603.
[5] Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet 2005;77(6):1044–60.
[6] Bemis EA, Demoruelle MK, Seifert JA, Polinski KJ, Weisman MH, Buckner JH, et al. Factors associated with progression to inflammatory arthritis in first-degree relatives of individuals with RA following autoantibody positive screening in a non-clinical setting. Ann Rheum Dis 2021;80(2):154–61.
[7] Frazzei G, van Vollenhoven RF, de Jong BA, Siegelaar SE, van Schaardenburg D. Preclinical autoimmune disease: a comparison of rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis and type 1 diabetes. Front Immunol 2022;13:899372.
[8] Rantapää-Dahlqvist S, de Jong BA, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis Rheum 2003;48(10):2741–9.
[9] Nielen MM, van Schaardenburg D, Reesink HW, van de Stadt RJ, van der Horst-Bruinsma IE, de Koning MHMT, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis Rheum 2004;50(2):380–6.
[10] Bos WH, Wolbink GJ, Boers M, Tijhuis GJ, de Vries N, van der Horst-Bruinsma IE, et al. Arthritis development in patients with arthralgia is strongly associated with anti-citrullinated protein antibody status: a prospective cohort study. Ann Rheum Dis 2010;69(3):490–4.

[11] Verheul MK, Böhringer S, van Delft MAM, Jones JD, Rigby WFC, Gan RW, et al. Triple positivity for anti-citrullinated protein autoantibodies, rheumatoid factor, and anti-carbamylated protein antibodies conferring high specificity for rheumatoid arthritis: implications for very early identification of at-risk individuals. Arthritis Rheumatol 2018;70(11):1721–31.
[12] Heliövaara M, Aho K, Aromaa A, Knekt P, Reunanen A. Smoking and risk of rheumatoid arthritis. J Rheumatol 1993;20(11):1830–5.
[13] Liu X, Tedeschi SK, Barbhaiya M, Leatherwood CL, Speyer CB, Lu B, et al. Impact and timing of smoking cessation on reducing risk of rheumatoid arthritis among women in the Nurses' Health Studies. Arthritis Care Res (Hoboken) 2019;71(7):914–24.
[14] Song H, Fang F, Tomasson G, Arnberg FK, Mataix-Cols D, Fernández de la Cruz L, et al. Association of stress-related disorders with subsequent autoimmune disease. JAMA 2018;319(23):2388–400.
[15] Hunt L, Hensor EM, Nam J, Burska AN, Parmar R, Emery P, et al. T cell subsets: an immunological biomarker to predict progression to clinical arthritis in ACPA-positive individuals. Ann Rheum Dis 2016;75(10):1884–9.
[16] Ponchel F, Burska AN, Hunt L, Gul H, Rabin T, Parmar R, et al. T-cell subset abnormalities predict progression along the inflammatory arthritis disease continuum: implications for management. Sci Rep 2020;10(1):3669.
[17] van de Stadt LA, Witte BI, Bos WH, van Schaardenburg D. A prediction rule for the development of arthritis in seropositive arthralgia patients. Ann Rheum Dis 2013;72(12):1920–6.
[18] van Steenbergen HW, Mangnus L, Reijnierse M, Huizinga TWJ, van der Helm-van Mil AHM. Clinical factors, anticitrullinated peptide antibodies and MRI-detected subclinical inflammation in relation to progression from clinically suspect arthralgia to arthritis. Ann Rheum Dis 2016;75(10):1824–30.
[19] Di Matteo A, Mankia K, Duquenne L, Cipolletta E, Wakefield RJ, Garcia-Montoya L, et al. Ultrasound erosions in the feet best predict progression to inflammatory arthritis in anti-CCP positive at-risk individuals without clinical synovitis. Ann Rheum Dis 2020;79(7):901–7.
[20] Frazzei G, Musters A, de Vries N, Tas SW, van Vollenhoven RF. Prevention of rheumatoid arthritis: a systematic literature review of preventive strategies in at-risk individuals. Autoimmun Rev 2023;22(1):103217.
[21] Gerlag DM, Safy M, Maijer KI, Tang MW, Tas SW, Starmans-Kool MJF, et al. Effects of B-cell directed therapy on the preclinical stage of rheumatoid arthritis: the PRAIRI study. Ann Rheum Dis 2019;78(2):179–85.
[22] Krijbolder DI, Verstappen M, van Dijk BT, Dakkak YJ, Burgers LE, Boer AC, et al. Intervention with methotrexate in patients with arthralgia at risk of rheumatoid arthritis to reduce the development of persistent arthritis and its disease burden (TREAT EARLIER): a randomised, double-blind, placebo-controlled, proof-of-concept trial. Lancet 2022;400(10348):283–94.
[23] Cope AP, Jasenecova M, Vasconcelos JC, Filer A, Raza K, Qureshi S, et al. Abatacept in individuals at high risk of rheumatoid arthritis (APIPPRA): a randomised, double-blind, multicentre, parallel, placebo-controlled, phase 2b clinical trial. Lancet 2024;403(10429):838–49.
[24] Rech J, Tascilar K, Hagen M, Kleyer A, Manger B, Schoenau V, et al. Abatacept inhibits inflammation and onset of rheumatoid arthritis in individuals at high risk (ARIAA): a randomised, international, multicentre, double-blind, placebo-controlled trial. Lancet 2024;403(10429):850–9.
[25] Ten Brinck RM, van Steenbergen HW, Mangnus L, Burgers LE, Reijnierse M, Huizinga TW, et al. Functional limitations in the phase of clinically suspect arthralgia are as serious as in early clinical arthritis; a longitudinal study. RMD Open 2017;3(1):e000419.
[26] Mankia K, Siddle HJ, Kerschbaumer A, Alpizar Rodriguez D, Catrina AI, Cañete JD, et al. EULAR points to consider for conducting clinical trials and observational studies in individuals at risk of rheumatoid arthritis. Ann Rheum Dis 2021;80(10):1286–98.
[27] Mankia K, Siddle H, Di Matteo A, Alpízar-Rodríguez D, Kerry J, Kerschbaumer A, et al. A core set of risk factors in individuals at risk of rheumatoid arthritis: a systematic literature review informing the EULAR points to consider for conducting clinical trials and observational studies in individuals at risk of rheumatoid arthritis. RMD Open 2021;7(3):e001768.
[28] Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569–81.
[29] Harnden K, Di Matteo A, Mankia K. When and how should we use imaging in individuals at risk of rheumatoid arthritis? Front Med (Lausanne) 2022;9:1058510.
[30] Ramos-Remus C, Castillo-Ortiz JD, Aguilar-Lozano L, Padilla-Ibarra J, Sandoval-Castro C, Vargas-Serafin CO, et al. Autoantibodies in prediction of the development of rheumatoid arthritis among healthy relatives of patients with the disease. Arthritis Rheumatol 2015;67(11):2837–44.
[31] Rakieh C, Nam JL, Hunt L, Hensor EM, Das S, Bissell LA, et al. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. Ann Rheum Dis 2015;74(9):1659–66.
[32] Sparks JA, Karlson EW. The roles of cigarette smoking and the lung in the transitions between phases of preclinical rheumatoid arthritis. Curr Rheumatol Rep 2016;18(3):15.

Ann Rheum Dis 84 (2025) 554–561

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Rheumatoid arthritis

Methotrexate continuation increases fracture risk in patients who sustained lower limb insufficiency fractures

Barbara Hauser[1,2,*], Andrew Merriman[1], Jonathan Foley[3], Janardhana Golla[1], Euan McRorie[1], Stuart H Ralston[1,2]

[1] *Rheumatic Diseases Unit, NHS Lothian, Western General Hospital, Edinburgh, UK*
[2] *Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Cancer, Western General Hospital, Edinburgh, UK*
[3] *Radiology Department, Royal Infirmary of Edinburgh, Edinburgh, UK*

ARTICLE INFO

Article history:
Received 22 September 2024
Received in revised form 4 January 2025
Accepted 6 January 2025

ABSTRACT

Objectives: Two recent case series and previous case reports have described methotrexate (MTX)-associated insufficiency fractures, so called methotrexate osteopathy (MTXO). Our aim was to assess whether the continuation of MTX after an insufficiency fracture impacts future fracture risk.

Methods: Retrospective single-centre case note review of patients who suffered MTXO insufficiency fractures. We assessed the occurrence of subsequent fractures and evaluated fracture healing in patients who either continued or discontinued MTX following the initial fracture.

Results: We identified 33 patients with characteristic MTXO lower limb insufficiency fractures. The mean MTX dose was 20 ± 5.9 mg weekly with average treatment duration of 10.7 ± 6.2 years. MTX was continued in 21 out of 32 patients following the initial insufficiency fracture. Almost all patients (95.2%) who continued methotrexate sustained either further insufficiency (67%) or major osteoporotic (33%) fractures. There were significantly fewer fractures (3 out of 11, 27.3%) in the group that stopped MTX after the initial insufficiency fracture (χ^2 = (1, N = 32) = 13.4; $P < .001$). A Kaplan–Meier analysis showed that significantly increased number of patients who continued MTX after the initial insufficiency fracture sustained a further fracture over time when compared to patients who stopped methotrexate ($P = .042$). Discontinuation of MTX was associated with greater clinical improvement in pain (77.8% vs 36.4%, $P = .036$) and weight-bearing capacity (71.4% vs 22.7%, $P = .030$) during fracture healing.

Conclusions: In patients with MTXO insufficiency fractures, continuation of MTX is associated with a high risk of further fracture. It is important to recognise such insufficiency fractures and stop MTX to minimise the future fracture risk.

*Correspondence to Dr. Barbara Hauser, Rheumatic Disease Unit, NHS Lothian, Western General Hospital, Edinburgh, UK.
E-mail address: Barbara.hauser@ed.ac.uk (B. Hauser).
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.01.047

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Methotrexate osteopathy (MTXO) has been previously identified in case reports, describing patients on long-term low-dose methotrexate developing insufficiency fractures in weight-bearing limbs.

WHAT THIS STUDY ADDS

- The study presents one of the largest case series to date on MTXO with detailed demographic and clinical data. It provides a novel analysis of the impact of methotrexate continuation on subsequent fracture risk. Methotrexate continuation after the initial fracture is associated with increased fracture rate compared to patients who stopped methotrexate.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- The study emphasises the importance of early diagnosis of methotrexate associated insufficiency fractures with advanced imaging such as magnetic resonance imaging.
- The diagnosis of MTXO should prompt re-evaluation of methotrexate use, as discontinuation may reduce the risk of future fractures and may allow better fracture healing.

INTRODUCTION

An association between methotrexate (MTX) and atypical fractures was first noted in the 1970s, when Ragab et al [1] observed that 5 children receiving high-dose MTX for acute leukaemia had developed insufficiency fractures of the lower limbs. These fractures were associated with 'metaphyseal osteoporosis' and a thickened zone of provisional calcification, an appearance normally associated with scurvy. Withdrawal of MTX resulted in improvement of symptoms and radiographic appearances. In 1984, Schwartz and Leonidas [2] described methotrexate osteopathy (MTXO) as a condition affecting children taking high-dose MTX for cancer, characterised by severe pain of the lower extremities, osteoporosis, and thickened zones of provisional calcification, with growth arrest lines and, sometimes, associated fractures. Since then, there have been a number of case reports describing MTXO in patients taking low-dose MTX for rheumatic disease [3–5].

More recently, further insights and a more detailed description of atypical stress fractures in patients taking long-term low-dose MTX has emerged from one larger case series with 34 patients [6], one further case series of 5 patients [7], and a systematic literature review on MTXO [8].

Rolvien et al [6] described a characteristic appearance of the stress fractures, with 'a unique band- or meander-shaped appearance along the growth plate'. These could occur in multiple, bilateral locations in weight-bearing limbs. The mechanism of these insufficiency fractures remains however poorly understood, and it is unclear whether MTX continuation impacts on future fracture risk.

We describe 33 patients on long-term MTX who developed multiple lower limb insufficiency fractures and the impact of MTX continuation or discontinuation on subsequent fracture rates.

METHODS

Case finding

Lower limb insufficiency fractures in patients with rheumatic diseases were identified by searching the magnetic resonance imaging (MRI) reports (requested by rheumatologists) for the term 'insufficiency fracture' in a large tertiary rheumatology centre from Jan 2019 to March 2021. Additionally, patients with insufficiency fractures were referred to the metabolic bone clinic as outlined in Figure 1. Patients with MTXO insufficiency fractures were included in the study until November 2022, ensuring a minimum follow-up period of 1 year. We conducted a retrospective review of identified cases using electronic health records and prospectively followed patients until December 2023 or until their death. The mean follow-up duration was 5.7 ± 4.6 years. We collected demographic data as well as clinical data, including diagnosis, disease activity, MTX treatment length and dose (last dose taken before first insufficiency fracture), concomitant antirheumatic treatments including current and previous glucocorticoid use, osteoporosis treatments, and underlying bone health parameters, including bone biochemistry, bone mineral density, fracture risk factors, and fracture risk as calculated using fracture risk assessment tool at the time of the first insufficiency fracture. Data on osteoporosis treatment prior and after the initial insufficiency fracture were obtained through case note review. Rheumatoid arthritis (RA) disease activity around the time of the first insufficiency fracture was assessed by recording either DAS 28 or inflammatory markers such as erythrocyte sedimentation rate (ESR) and/or C-reactive protein (CRP) performed at the time. The date of the diagnostic imaging of the incident (first) insufficiency fracture was recorded and through a search of electronic healthcare and imaging records we identified subsequent insufficiency, or major osteoporotic fractures. Vertebral fractures were typically assessed through thoracolumbar spine radiographs. These radiographs were only performed in patients with back pain or significant height loss (>4 cm).

Imaging

The majority of insufficiency fracture cases were identified and confirmed with MRI, computed tomography (CT), or 99Tc radionuclide bone scan imaging. Only a one-third of cases were diagnosed by plain radiographs. All images were reviewed and reported by radiologists as per routine clinical care. Bone mineral density (BMD) measurements were made by dual-energy x-ray absorptiometry at the lumbar spine (L1–L4) and femoral neck using a QDR 4500 osteodensitometer (Hologic).

Fracture definition

According to previous reports [8], MTXO insufficiency fractures were diagnosed if the following characteristics were fulfilled: (1) prolonged methotrexate use; (2) typical appearance of incomplete fracture with band-like fracture lines along the growth plate; (3) anatomical location, including metaphysis of distal or proximal tibia, distal femur, calcaneus or talus; and (4) low impact fractures. Concomitant glucocorticoid use was not considered an exclusion criterion, as the described insufficiency fractures exhibit a distinct fracture pattern, and previous reports on MTXO included patients with recent or ongoing glucocorticoid use [6,7]. Two cases with typical MRI features of MTXO insufficiency fractures are shown in Figure 2A,B.

Fracture healing assessment

Fracture healing was evaluated using both radiological and clinical assessments. A radiologist compared follow-up imaging

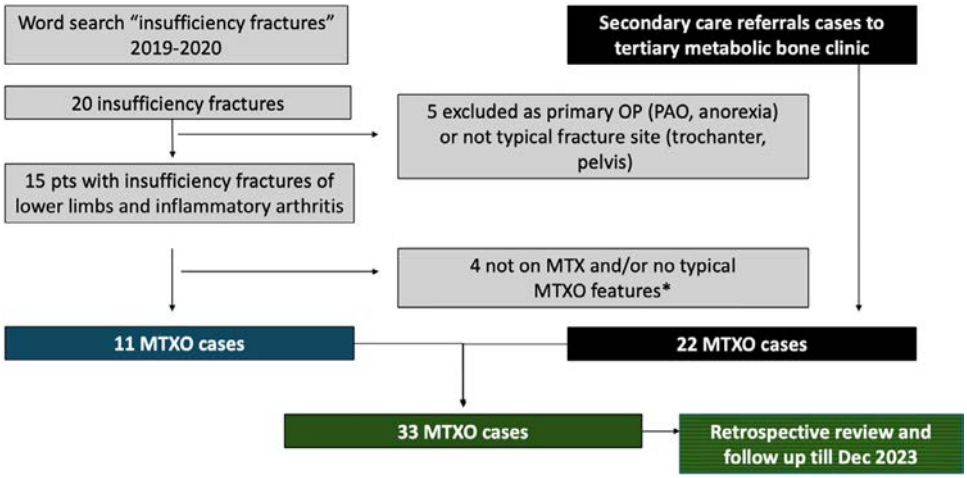


Figure 1. Flow diagram of case collection of patients with suspected MTX osteopathy. MTX, methotrexate; MTXO, methotrexate osteopathy; OP, osteoporosis; PAO, pregnancy-associated osteoporosis; pts, patients.

with the original images of the first insufficiency fracture, focusing on parameters such as the presence or absence of callus formation, persistence of the fracture line, and evidence of radiological fracture resolution. Clinical data were extracted from routine evaluations conducted primarily by rheumatologists or orthopaedic specialists. These assessments included improvements in pain, the ability to bear weight, and tenderness on palpation at the fracture site.

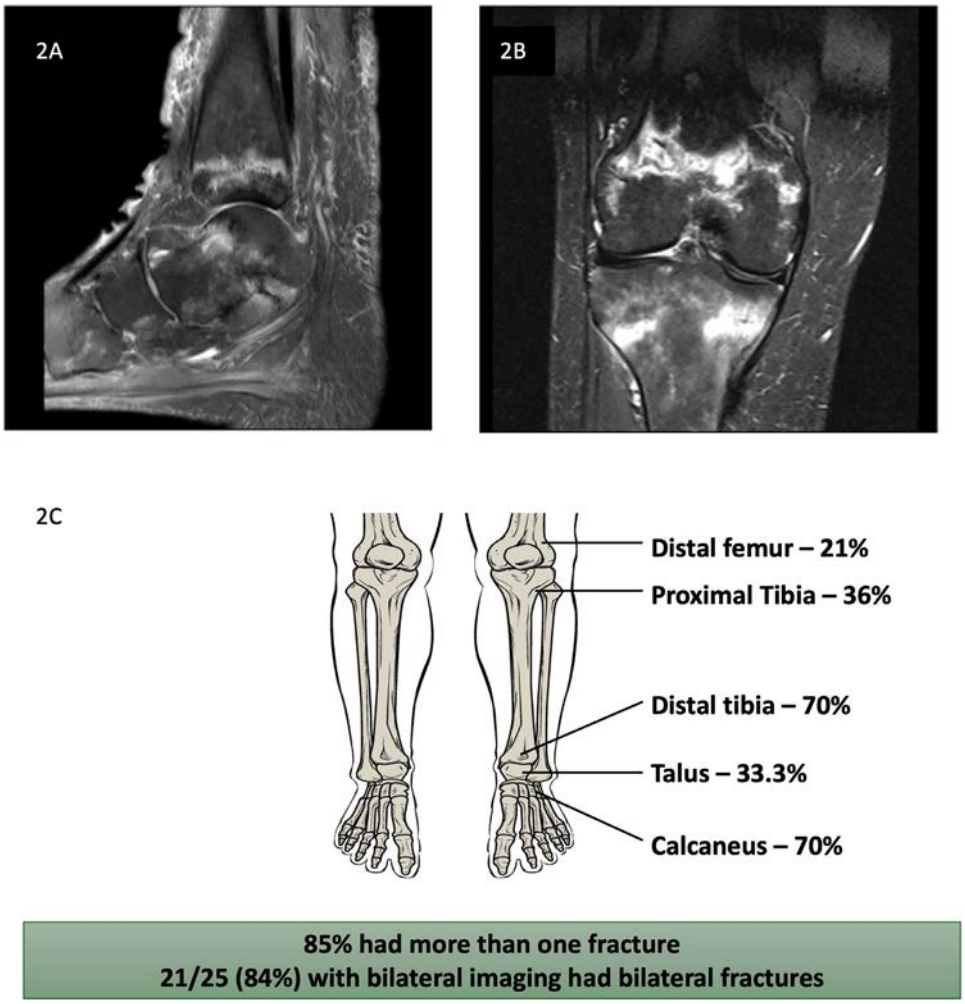


Figure 2. (A and B) MTXO fractures. (A) MRI of the left ankle (sagittal Proton Density Fat Saturation (PDFS) sequence), showing an insufficiency fracture of the distal tibial metaphysis and a sagittal oblique fracture of the medial talus. (B) MRI T1 Turbo Inversion Recovery Magnitude (TIRM) sequence, highlighting bone oedema at the metaphyseal closure line of the femur and metaphyseal closure line of tibial condyles, denoting insufficiency fracturing. (C) Distribution and frequency of insufficiency fractures. MRI, magnetic resonance imaging.

MTXO continuation versus discontinuation group

The date of imaging which confirmed features of MTXO insufficiency fracture (as outlined in the fracture definition above) was recorded. The responsible healthcare provider decided on continuation of MTX depending on whether MTX was thought to be implicated as contributing factor to the fracture or not. Further fractures were identified, reviewed and recorded through review of the electronic health records and the national electronic imaging database Picture Archiving Communication System. Clinical and radiographic features of fracture healing were recorded in both groups.

Statistical analysis

Statistical analyses were performed using SPSS version 24.0 (IBM Corp). Data were described as the total number (*n*), percentage (%), mean with SD in parentheses or median with IQR or minimum and maximum values. Some missing data occurred due the retrospective collection of data, which was applicable to inflammatory markers (CRP and ESR), DAS 28, BMD, and bone biochemistry markers. Missing data points are shown in the respective tables (Tables 1 and 2). Independent student *t* test was used for between-group comparisons of continuous data if normally distributed and Mann-Whitney U test for comparison of continuous data if not normally distributed. Chi square test (χ^2) was used to evaluate differences between categorical variables. Kaplan–Meier analysis assessed rate and time of fractures which occurred after the first MTXO insufficiency fracture in patients who continued or stopped MTX. Statistical significance was assessed with Cox regression with fracture (yes/no) as time-dependent covariate. Statistical significance was defined as a *P* value <.05 for all analyses.

Ethics

All patient data were evaluated in a retrospective and anonymised design. Information about the MTXO population was obtained through an audit of data that had been collected during routine clinical practice; hence, ethical approval was not required.

RESULTS

Demographics

We identified 33 patients with characteristic MTXO lower limb insufficiency fractures. The mean MTX dose was 20 ± 5.9 mg weekly with an average treatment duration of 10.7 ± 6.2 years. The shortest duration of continuous MTX intake in this group was 20 months with a cumulative MTX dose of 2220 mg administered subcutaneously. The mean age of the cohort was 68 ± 9 years, with a predominant female population (96%) and a high prevalence of RA (88%). Patients with other conditions such as psoriatic arthritis, psoriasis, and undifferentiated connective tissue disease were also affected. The average body mass index was 25.6 ± 5.1 kg/m². The demographics and clinical characteristics are similar to the largest case series published by Rolvien et al [6], except that none of our patients were on continuous long-term oral glucocorticoid (GC) treatment at the time of the first insufficiency fracture. Additionally, similar to other large case series, we observed a delay in diagnosing MTX insufficiency fractures; however, in our cohort, the delay was 7.3 months compared to 17.4 months in the other series. On average patients, were treated for over 10 years when they presented with the first MTXO fracture, and there was no significant difference between subcutaneous or oral MTX administration in respect to MTX duration.

Table 1
Demographic and clinical details of patients with MTXO fractures (at the time of first insufficiency fracture)

Demographics	MTXO cohort (33)
Female (%)	32 (97%)
Age (y)	67.3 ± 8.9
BMI (kg/m^2)	25.7 ± 4.9
Underlying medical diagnosis	
RA (23/28 seropositive)	28 (85%)
Other (3 psoriatic arthritis, 1 psoriasis, 1 UCTD)	5 (15%)
Antirheumatic treatments	
Current glucocorticoid treatment (including 3 mo before first IF)	4/31 (13%)
Systemic GC (course of oral or im GC)	3/31 (9.7%)
Cumulative prednisolone dose of systemic GC (mg)	*203.3 + 138.0*
Intra-articular GC injection	2/31 (6.5%)
Cumulative prednisolone dose of intra-articular injections (mg)	*125 + 35.4*
Long-term continuous GC intake	0 (0%)
Past glucocorticoid treatment (including 24 to 3 mo before first IF)	10/31 (32%)
Systemic GC (course of oral GC or im GC)	5/31
Cumulative prednisolone dose of systemic GC (mg)	*303.0 ± 138.4*
Intra-articular GC injection	5/31
Cumulative prednisolone dose of ia GC (mg)	*80.0 ± 67.1*
Overall weekly MTX dose (mg)	20.1 ± 5.6
Mean MTX dose (mg/kg)	0.31 ± 0.11
MTX oral administration	16/32 (50%)
Mean dose (oral MTX) (mg/wk)	17.2 ± 6.6
MTX subcutaneous administration	16/32 (50%)
Mean dose (sc MTX) (mg/wk)	22.2 ± 4.4
MTX treatment length (y)	10.7 ± 6.2
Folic acid (5 mg/wk)	28/32 (88%)
Other folic acid dose (30 mg/wk; 15 mg/wk)	3 (9%); 1 (3%)
MTX monotherapy	22/33 (67%)
Concomitant conventional DMARDS in addition to MTX	6 /33 (18%)
Hydroxychloroquine (4)	
Leflunomide (1)	
Sulfasalazine (1)	
Concomitant biologic DMARDS (TNFi)	5/33 (15%)
Certolizumab (3), adalimumab (1), etanercept (1)	
Symptom onset to diagnosis of MTXO (mo)	7.3 ± 8.2
MRI/CT/NM bone scan diagnosis of first insufficiency fracture	23 (70%)

BMI, body mass index; CT, computer tomography; F, female, GC, glucocorticoid; ia, intra-articular; IF, insufficiency fracture; im, intramuscular; MRI, magnet resonance imaging; MTX, methotrexate; NM, nuclear medicine 99Tc radionuclide bone scan; RA, rheumatoid arthritis; sc, subcutaneous; UCTD, undifferentiated connective tissue disease; DMARD, disease modifying antirheumatic drug; TNFi, Tumor necrosis factor inhibitor.

Bone health characteristics

The most common insufficiency fracture sites were the metaphysis of the distal tibia (70%) and calcaneus (70%), followed by the proximal tibia, talus, and distal femur (see Fig 2C). Notably, 84% of the patients experienced more than 1 fracture, and 21 out of 25 patients who had imaging of both limbs were found to have bilateral fractures. Advanced imaging such as MRI, CT or 99Tc radionuclide bone scan imaging was required for 70% of the patients to confirm the diagnosis, with an average delay of 7 ± 8 months between initial symptoms and diagnostic confirmation. Osteoporosis was present in 75% of the patients, with an average spine T score of −2.3 ± 1.0 and a femoral neck T score of −2.6 ± 0.9. The incidence of previous vertebral fractures was 16.9%. Biochemical markers were on average within normal

Table 2
Clinical characteristics: bone mineral density, osteoporosis medication, and biochemical markers

Clinical characteristics	Lothian cohort (33)
Osteoporosis as per DXA (T score <−2.5)	25/32 (75%)
BMD T score spine	−2.3 ± 1.0
BMD T score femoral neck	−2.6 ± 0.9
Previous vertebral fractures	8 (16.9%)
Calcium (2.2-2.6 mmol/L)	2.4 ± 0.1
Vitamin D (>50 nmol/L adequate)	48.6 ± 27.0
Alkaline phosphatase (40-125 U/L)	101.5 ± 23.4
ESR (mm/hr)	21.4 ± 22.9
DAS 28 [n = 9]	3.2 ± 1.3
OP treatment prior to MTXO# (n, %)	8/32 (24.2%)
Alendronic acid, 70 mg orally weekly	6 (19%)
Risedronic acid, 35 mg orally weekly	2 (6%)
Mean OP treatment time prior MTXO# (y)	3.6 ± 1.5
OP treatment started after MTXO# (n, %)	28/31 (90.3%)
Alendronic acid, 70 mg orally weekly	16 (52%)
Risedronic acid, 35 mg orally weekly	1 (3%)
Zoledronic acid, 5 mg intravenous yearly	7 (23%)
Teriparatide, 20 μg subcutaneous daily	2 (6%)
Romosozumab, 210 mg subcutaneous monthly	2 (6%)

BMD, bone mineral density; DAS28, disease activity score including 28 joints; DXA, dual-energy x-ray absorptiometry; ESR, erythrocyte sedimentation rate; MTX, methotrexate; MTXO#, methotrexate osteopathy insufficiency fractures; OP, osteoporosis.

range as shown in Table 2. Mean ESR was 21.4 ± 22.9 mm/hr, and the average DAS28 score for a subset of patients (3.2 ± 1.3) indicated mild to moderate disease activity.

Impact of methotrexate continuation

Following the initial insufficiency fracture, 21 out of 32 patients continued MTX, and nearly all of these patients (95.2%) sustained further insufficiency (67%) or major osteoporotic (33%) fractures. For 1 patient, we could not obtain the information whether they continued MTX or not as we were unable to obtain their medical records. For patients who discontinued MTX, only 4 out of 11 patients (36%) experienced subsequent fractures. Further details on both groups are shown in Table 3. In follow-up assessments of fracture healing, patients who discontinued MTX were more likely to report clinical improvement compared to those who continued the medication. Specifically, a higher proportion of patients who discontinued MTX reported improvement in pain (77.8%) and the ability to fully weight bear (71.45%) compared to those who continued MTX (36.4% and 22.7%, respectively). Despite these clinical differences, radiological signs of fracture persisted in the majority of patients (77.8%), with no significant difference in radiological resolution observed between the 2 groups.

A Kaplan–Meier analysis, shown in Figure 3, demonstrated that patients who continued MTX had a significantly higher incidence of further fractures over time compared to those who stopped MTX (P = .042).

DISCUSSION

This study describes one of the largest cohort of patients (n = 33) with MTXO insufficiency fractures and, uniquely, provides data on subsequent fracture history depending on whether or not MTX was continued or stopped. The patient characteristics, stress fracture location, and the distinctive appearance of the insufficiency fractures are consistent with previous reports and case series of MTXO. Also, in keeping with previous reports, our cohort demonstrated a high incidence of bilateral fractures and a significant delay in diagnosis.

A significant contribution of our study is the analysis of the impact of MTX continuation after the initial insufficiency fracture on subsequent fracture risk. Although previous studies have identified the occurrence of MTXO, our research is the first to provide evidence that continuing MTX therapy after an initial insufficiency fracture not only significantly increases the risk of further insufficiency fractures there was also a numerical increase in major osteoporotic fractures in patients who continued MTX. This novel insight underscores the need to re-evaluate MTX therapy in patients presenting with MTXO.

Our study also highlights the difficulty in diagnosing MTXO. The plain radiographic findings at presentation can be subtle, and detailed imaging such as MRI or CT imaging is often necessary for the diagnosis to be made [6–8].

Additional confounding factors that may lead to a delay in diagnosis include physician awareness of the diagnosis of MTXO; the potential misinterpretation of symptoms as being

Table 3
Patients who stopped MTX versus patients who continued

Characteristics	MTX continuation 21 (66%)	MTX stopped 11 (34%)	P value
Age (y)	66.6 ± 9.9	68.2 ± 7.2	.677
Mean follow-up time (mo)	77.9 ± 48.6	53.8 ± 66.4	.303
FRAX 10-y-MOP (n = 29)	25.5 ± 9.9	19.4 ± 8.0	.114
FRAX 10 y-Hip (n = 29)	10.5 ± 7.7	6.7 ± 5.6	.155
Spine T score (n = 29)	−2.4 ± 1.0	−2.2 ± 1.1	.690
Hip T score (n = 28)	−2.4 ± 1.0	−2.1 ± 0.5	.238
Fractures			
Time to subsequent fractures (mo) (insufficiency and MOP fractures)	18.7 ± 22.3	65.4 ± 113.1	.081
Total subsequent fractures (insufficiency and MOP fractures)	20 (95%)	4 (36%)	.001
Insufficiency fractures	13 (65)%	2 (50%)	
MOP fractures	7(33%)	2 (50%)	
Radiological and clinical signs of fracture healing			
Radiological fracture resolution (n = 28)	3 (18.8%)	3 (27.3%)	.754
Callus formation (n = 30)	5 (26.3%)	1 (9.1%)	.372
Ongoing fracture line visibility (n = 29)	11 (61.1%)	6 (54.5%)	.728
Pain improvement (n = 31)	8 (36.4 %)	7 (77.8 %)	.036
Tenderness to touch (n = 24)	16 (94.1)	4 (57.1)	.059
Weight-bearing tolerance (n = 29)	5 (22.7)	5 (71.4)	.030

FRAX, fracture risk assessment tool; MOP, major osteoporotic fracture.

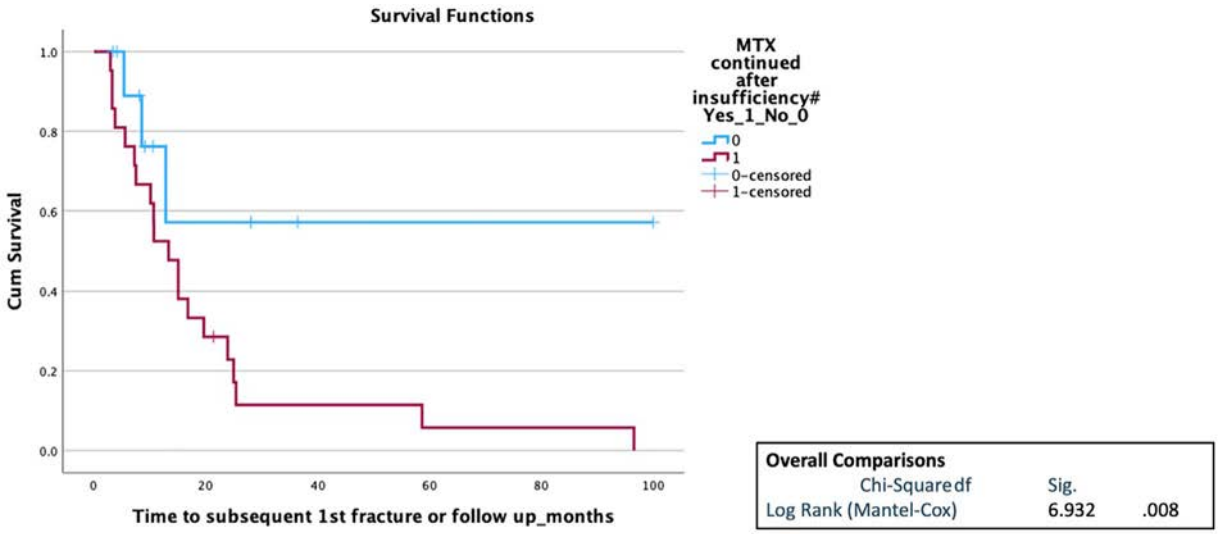


Figure 3. Kaplan–Meier curve for MTX continuation and discontinuation after an initial insufficiency fracture. The light blue line depicts the patient group who have stopped MTX after the initial insufficiency fracture. The purple line depicts the patient group who continued MTX. MTX, methotrexate.

due to joint disease, particularly in those patients for whom MTX is being prescribed to treat inflammatory arthritis; and the assumption that the fracture might be due to underlying osteoporosis, which often coexists with MTX use.

Our study provides details on route of MTX administration and concomitant folic acid medication. The route of administration of MTX is unlikely to be relevant as half of MTXO patients were administering MTX subcutaneously and half were taking the oral formulation. All patients were prescribed concomitant folic acid with 5 mg/week being the most frequently used dose. Importantly the vast majority of above described patients have not received recent GCs, and 2 out of 3 patients received intra-articular steroid injections to joints which were most likely painful due to developing insufficiency fractures. Hence, it is unlikely that these insufficiency fractures represent GC-induced osteoporosis. Based on our experience and previous reports [6,7] as well as the observation that the described insufficiency fractures differ from typical GC-induced osteoporotic fractures, such as vertebral, hip, pelvis or rib fractures [9], systemic GC use is currently not considered to be an exclusion criterion in diagnosing MTXO. However, given the well documented detrimental effects of high cumulative GC doses on bone turnover, including profound suppression of bone formation and negative impacts on bone quality [10,11], the role of GCs in the development and treatment of MTXO insufficiency fractures remains to be fully elucidated.

The effect of MTX on fracture risk is likely idiosyncratic as only a small proportion of individuals receiving MTX for inflammatory rheumatic diseases experience this problem. In addition, MTX use has not been associated to accelerated bone loss or to increased osteoporotic fracture risk. A small case control study has shown no difference in BMD between RA or psoriatic arthritis patients treated with MTX over 6 years or controls [12]. A more extensive and recent dataset from the Women's Health Initiative observational study followed 1201 women with RA for approximately 6 years. The study found no significant difference in the self-reported clinical fracture rates between patients who were treated with sulfasalazine or MTX [13]. There is no evidence to suggest that methotrexate use increases the risk of major osteoporotic fractures [13]. However, we hypothesise that the observed numerical difference in our cohort may be attributed to the fact that patients with multiple, recurrent insufficiency fractures and delayed healing are more likely to experience disuse bone and muscle loss. This, in addition to the elevated fracture risk associated with the underlying rheumatic disease and other potential risk factors, may contribute to an increased risk of major osteoporotic fractures [13].

Despite multiple reported cases and the more detailed characterisation of MTXO over the recent years, the question of whether MTX truly impacts on fracture risk and the development of insufficiency fractures has been raised [14,15]. Guidelines outlining or including the efficacy and safety of MTX for the use of RA or psoriatic arthritis have highlighted MTX as gold standard treatment but have so far omitted to mention MTXO as a potential serious side effect [16–18].

Furthermore, our study supports the hypothesis that continuous MTX therapy may delay the healing of insufficiency fractures [19,20]. Patients who discontinued MTX therapy more frequently reported improvements in pain and demonstrated better ability to fully bear weight during follow-up assessments. Although no significant differences were observed in radiological signs of fracture healing, this may be due to the lack of consecutive, detailed imaging in most patients.

Although the above findings provide important insights, several limitations must be acknowledged. One significant limitation of this study is its retrospective nature, hence limited by the quality and completeness of existing records. Several studies have shown that RA is associated with increased fracture risk [21,22] and that high disease activity is a particular important risk factor [23,24]. In our RA cohort, the disease activity seemed to be low to moderate; however, disease activity scores such as DAS28 were not consistently recorded.

Additionally, the identification of insufficiency fractures in radiology reports was limited to MRI scans specifically requested by rheumatologists, thereby fractures investigated and managed by orthopaedic teams may not have been included. We suggest further epidemiological studies using large scale electronic health record data sets to define the prevalence of MTXO. Additional areas to investigate include the influence of specific risk factors, such as genetic predisposition, on susceptibility to MTXO. Genetic variations associated with bone

turnover, fracture risk, or drug metabolism might increase an individual's likelihood of developing this condition. To examine this, we propose utilising a candidate gene approach, comparing MTXO patients with matched controls.

Another limitation is the inconsistency in imaging modalities used to assess further fractures. Detailed imaging techniques such as MRI, CT, and radionuclide bone scans were required for the majority of patients to confirm insufficiency fractures. However, imaging of the contralateral side was not consistently performed, which could have resulted in underreporting of additional fractures. Additionally, the absence of systematic follow-up imaging to monitor fracture healing limits our ability to fully understand the progression and resolution of MTXO-related fractures. We recommend that future studies should follow patients with MTXO insufficiency fractures systematically with regular clinical exams, regular imaging of the contralateral side and sequential follow-up for imaging of the insufficiency fractures to assess fracture healing.

Despite these limitations, our study has several strengths. It is one of the largest case series to date examining MTXO, providing detailed demographic and clinical data. Furthermore, the study offers a novel analysis of the impact of MTX continuation on subsequent fracture risk, which provides additional evidence for the entity of MTXO. The study highlights the association between continued MTX use following an insufficiency fracture and an increased risk of subsequent fractures. Early recognition of MTXO using advanced imaging is critical to prevent further fractures. Discontinuation of MTX after an initial insufficiency fracture reduces fracture risk and likely improves fracture healing. These findings underscore the need to re-evaluate MTX therapy in patients with MTXO. Future research should focus on identifying genetic and clinical risk factors for MTXO and conducting larger studies to confirm our findings and guide clinical management.

Competing interests

BH reports personal funding from UCB, Amgen, EffrX, Thornton Ross and Eli Lilly and funding to her institution from Kyowa Kirin, and UCB, outside the submitted work. SHR reports funding to his institution from Kyowa Kirin, and UCB, outside the submitted work. AM, JF, JG, and EM declare no conflicts of interest.

Acknowledgements

We thank the rheumatology staff at the Western General Hospital for the help and support in completing this project.

Contributors

Conceptualisation: BH, AM, SHR; Methodology: BH, AM, SHR; Investigation: BH, AM; Writing—original draft: BH; Writing—review and editing: BH, AM, JF, JG, EM, SHR; Supervision: BH. BH was responsible for the overall content as the guarantor. All authors reviewed and approved of the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Patient consent for publication

Not applicable

Data availability statement

Data are available upon reasonable request.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Orcid

Barbara Hauser: http://orcid.org/0000-0003-2693-1957
Andrew Merriman: http://orcid.org/0009-0006-5034-5228
Jonathan Foley: http://orcid.org/0000-0002-5877-4768
Janardhana Golla: http://orcid.org/0009-0006-4610-9388
Euan McRorie: http://orcid.org/0009-0008-7339-1686
Stuart H Ralston: http://orcid.org/0000-0002-2804-7586

REFERENCES

[1] Ragab AH, Frech RS, Vietti TJ. Osteoporotic fractures secondary to methotrexate therapy of acute leukemia in remission. Cancer 1970;25(3):580–5. doi: 10.1002/1097-0142(197003)25:3<580::AID-CNCR2820250313>3.0.CO;2-M.
[2] Schwartz AM, Leonidas JC. Methotrexate osteopathy. Skeletal Radiol 1984;11(1):13–6. doi: 10.1007/BF00361126.
[3] Ansell G, Evans S, Jackson CT, Lewis-Jones S. Cytotoxic drugs for non-neoplastic disease. Br Med J (Clin Res Ed) 1983;287(6394):762. doi: 10.1136/bmj.287.6394.762-a.
[4] Singwe M, Le Gars L, Karneff A, Prier A, Kaplan G. Multiple stress fractures in a scleroderma patient on methotrexate therapy. Rev Rhum Engl Ed 1998;65(7-9):508–10.
[5] Wijnands M, Burgers A. Stress fracture in long term methotrexate treatment for psoriatic arthritis. Ann Rheum Dis 2001;60(8):736–9. doi: 10.1136/ard.60.8.736.
[6] Rolvien T, Jandl NM, Stürznickel J, Beil FT, Kötter I, Oheim R, et al. Clinical and radiological characterization of patients with immobilizing and progressive stress fractures in methotrexate osteopathy. Calcif Tissue Int 2021;108(2):219–30. doi: 10.1007/s00223-020-00765-5.
[7] Robin F, Cadiou S, Albert JD, Bart G, Coiffier G, Guggenbuhl P. Methotrexate osteopathy: five cases and systematic literature review. Osteoporos Int 2021 Feb;32(2):225–32. doi: 10.1007/s00198-020-05664-x.
[8] Ruffer N, Krusche M, Beil FT, Amling M, Kötter I, Rolvien T. Clinical features of methotrexate osteopathy in rheumatic musculoskeletal disease: a systematic review. Semin Arthritis Rheum 2022;52:151952. doi: 10.1016/j.semarthrit.2022.151952.
[9] Compston J. Glucocorticoid-induced osteoporosis: an update. Endocrine 2018;61(1):7–16. doi: 10.1007/s12020-018-1588-2.
[10] Hartmann K, Koenen M, Schauer S, Wittig-Blaich S, Ahmad M, Baschant U, et al. Molecular actions of glucocorticoids in cartilage and bone during health, disease, and steroid therapy. Physiol Rev 2016;96(2):409–47. doi: 10.1152/physrev.00011.2015.
[11] Lane NE. Glucocorticoid-induced osteoporosis: new insights into the pathophysiology and treatments. Curr Osteoporos Rep 2019;17(1):1–7. doi: 10.1007/s11914-019-00498-x.
[12] Cranney AB, McKendry RJ, Wells GA, Ooi DS, Kanigsberg ND, Kraag GR, et al. The effect of low dose methotrexate on bone density. J Rheumatol 2001;28(11):2395–9.
[13] Carbone L, Vasan S, Elam R, Gupta S, Tolaymat O, Crandall C, et al. The association of methotrexate, sulfasalazine, and hydroxychloroquine use with fracture in postmenopausal women with rheumatoid arthritis: findings from the women's health initiative. JBMR Plus 2020;4(10):e10393. doi: 10.1002/jbm4.10393.
[14] Mönig H, Ullrich S. [Does MTX-induced osteoporosis exist?]. Z Rheumatol 2020;79(9):898–901. doi: 10.1007/s00393-020-00855-7.
[15] Maenaut K, Westhovens R, Dequeker J. Methotrexate osteopathy, does it exist? J Rheumatol 1996;23(12):2156–9.

[16] Valerio V, Kwok M, Loewen H, Winkler J, Mody GM, Scuccimarri R, et al. Systematic review of recommendations on the use of methotrexate in rheumatoid arthritis. Clin Rheumatol 2021;40(4):1259–71. doi: 10.1007/s10067-020-05363-2.

[17] Kerschbaumer A, Smolen JS, Ferreira RJO, Bertheussen H, Baraliakos X, Aletaha D, et al. Efficacy and safety of pharmacological treatment of psoriatic arthritis: a systematic literature research informing the 2023 update of the EULAR recommendations for the management of psoriatic arthritis. Ann Rheum Dis 2024;83(6):760–74. doi: 10.1136/ard-2024-225534.

[18] Smolen JS, Landewé RBM, Bergstra SA, Kerschbaumer A, Sepriano A, Aletaha D, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis 2023;82(1):3–18. doi: 10.1136/ard-2022-223356.

[19] Singh D, Hesse N, Skapenko A, Schulze-Koops H. A rare side effect of methotrexate therapy: drug-induced osteopathy with multiple fractures of the lower limb. RMD Open 2023;9(1):e002982. doi: 10.1136/rmdopen-2023-002982.

[20] Rolvien T, Creutzfeldt AM, Lohse AW. Amling M. Stress fractures in systemic lupus erythematosus after long-term MTX use successfully treated by MTX discontinuation and individualized bone-specific therapy. Lupus 2019;28(6):790–3. doi: 10.1177/0961203319841434.

[21] Xue AL, Wu SY, Jiang L, Feng AM, Guo HF, Zhao P. Bone fracture risk in patients with rheumatoid arthritis: a meta-analysis. Medicine (Baltimore) 2017;96(36):e6983. doi: 10.1097/MD.0000000000006983.

[22] Hauser B, Riches PL, Wilson JF, Horne AE, Ralston SH. Prevalence and clinical prediction of osteoporosis in a contemporary cohort of patients with rheumatoid arthritis. Rheumatology (Oxford) 2014;53(10):1759–66. doi: 10.1093/rheumatology/keu162.

[23] Haugeberg G, Ørstavik RE, Uhlig T, Falch JA, Halse JI, Kvien TK. Clinical decision rules in rheumatoid arthritis: do they identify patients at high risk for osteoporosis? Testing clinical criteria in a population based cohort of patients with rheumatoid arthritis recruited from the Oslo Rheumatoid Arthritis Register. Ann Rheum Dis 2002;61(12):1085–9. doi: 10.1136/ard.61.12.1085.

[24] Tong JJ, Xu SQ, Zong HX, Pan MJ, Teng YZ, Xu JH. Prevalence and risk factors associated with vertebral osteoporotic fractures in patients with rheumatoid arthritis. Clin Rheumatol 2020;39(2):357–64. doi: 10.1007/s10067-019-04787-9.

Ann Rheum Dis 84 (2025) 562–574

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Rheumatoid arthritis

Dynamics of the gut microbiome in individuals at risk of rheumatoid arthritis: a cross-sectional and longitudinal observational study

Christopher M. Rooney[1,*], Ian B. Jeffery[2], Kulveer Mankia[3,4], Mark H. Wilcox[1], Paul Emery[3,4]

[1] *Leeds Institute of Medical Research, University of Leeds, Leeds, UK*
[2] *4D Pharma PLC, Cork, Ireland*
[3] *University of Leeds, Leeds Institute of Rheumatic and Musculoskeletal Medicine, Leeds, UK*
[4] *NIHR Leeds Musculoskeletal Biomedical Research Centre, Leeds, UK*

ARTICLE INFO

Article history:
Received 4 July 2024
Accepted 3 October 2024

Keywords:
Rheumatoid Arthritis
Anti-Citrullinated Protein Antibodies
Risk Factors

ABSTRACT

Objectives: This work aimed to resolve the conflicting reports on Prevotellaceae abundance in the development of rheumatoid arthritis (RA) and to observe structural, functional and temporal changes in the gut microbiome in RA progressors versus non-progressors.

Methods: Individuals at risk of RA were defined by the presence of anticyclic citrullinated protein (anti-CCP) antibodies and new musculoskeletal symptoms without clinical synovitis. Baseline sampling included 124 participants (30 progressed to RA), with longitudinal sampling of 19 participants (5 progressed to RA) over 15 months at five timepoints. Gut microbiome taxonomic alterations were investigated using 16S rRNA amplicon sequencing and confirmed with shotgun metagenomic DNA sequencing on 49 samples.

Results: At baseline, CCP+ at risk progressors showed significant differences in Prevotellaceae abundance compared with non-progressors, contingent on intrinsic RA risk factors and time to progression. Longitudinal sampling revealed gut microbiome instability in progressors 10 months before RA onset, a phenomenon absent in non-progressors. This may indicate a late microbial shift before RA onset, with Prevotellaceae contributing but not dominating these changes. Structural changes in the gut microbiome during arthritis development were associated with increased amino acid metabolism.

Conclusion: These data suggest conflicting reports on Prevotellaceae overabundance are likely due to sampling within a heterogeneous population along a dynamic disease spectrum, with certain Prevotellaceae strains/clades possibly contributing to the establishment and/or progression of RA. Gut microbiome changes in RA may appear at the transition to clinical arthritis as a late manifestation, and it remains unclear whether they represent a primary or secondary phenomenon.

*Correspondence to Dr. Christopher M. Rooney.
E-mail address: c.rooney@leeds.ac.uk (C.M. Rooney).
Social media: @DrCRooney (C. M. Rooney)
Handling editor Josef S. Smolen

https://doi.org/10.1136/ard-2024-226362

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Individuals with rheumatoid arthritis (RA) and those at risk possess distinct gut microbiomes when compared with healthy controls; however, detailed insights into the temporal and compositional changes in the gut microbiome as individuals progress to RA are lacking.
- Prevotellaceae, and particularly *Prevotella copri*, have been inconsistently associated with RA development, and while early studies showed an overabundance more recent cohorts that question these observations have emerged.

WHAT THIS STUDY ADDS

- We identified a phenomenon where specific strains/clades of Prevotellaceae are enriched, while others are depleted, depending on an individual's risk profile and time to arthritis progression.
- Anticyclic citrullinated protein antibody-positive individuals at risk of RA are characterised by a low-diversity gut microbiome that may become more unstable in the 10 months preceding the transition from preclinical RA to established disease (ie, the development of clinical arthritis).
- This instability is marked by accumulation of multiple bacteria, including but not limited to Prevotellaceae.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- These findings illustrate the dynamic nature of the gut microbiome, reflecting an individual's underlying risk profile and transition from being at risk to developing clinical arthritis.
- This evolving understanding could significantly impact approaches to RA prediction, prevention and personalised treatment strategies, emphasising the need for further research into microbiome-based diagnostics and therapeutics.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic autoimmune disorder affecting over half a million individuals in the UK. The hallmark of RA is progressive joint disease, with potential for systemic involvement. Understanding the RA disease spectrum with recognition of at risk individuals has propelled RA research into prevention strategies. The generation of IgA class anticitrullinated protein antibodies (ACPAs) in individuals at risk of RA, [1] combined with epidemiological links with smoking [2] and periodontal disease [3], points to a mucosal origin of inflammation. The mucosal origin hypothesis proposes that localised inflammation at mucosal sites can initiate a broader immune response [1], via T cell activation and a subsequent inflammatory cytokine cascade, leading to B cell antibody production. Supporting this, an immunoglobulin class switch from IgA ACPAs to IgG ACPAs [4,5] indicates potential triggering of systemic autoimmunity by diverse antigenic stimuli at mucosal sites. This shift, accompanied with broadening of antibody targets [6,7], suggests that mucosal barrier deterioration and the ensuing spread of an IgG ACPA response might be more significant in the initial stages of RA than the loss of tolerance to self-antigens.

Profiling of the gut microbiome in individuals at risk of RA and people diagnosed with RA consistently demonstrates a dysbiotic microbiome when compared with healthy controls [8–21]. However, there remains little consensus on the bacterial constituent members of an RA-related dysbiosis. Subsequently, a variety of gut bacteria have been implicated as a potential impetus in the development of RA, none more so than *Prevotella copri*. *Prevotella* species have been demonstrated to be overabundant in new-onset rheumatoid arthritis (NORA) [9,11,22], in at risk individuals [8] and especially those with genetic risk [10]. Their abundance decreases after disease-modifying antirheumatic drug (DMARD) therapy, with reversion to a eubiotic state on treatment [22]. Furthermore, mouse models support a role for Prevotellaceae strains derived from patients with RA in RA development [23]. However, Prevotellaceae overabundance does not appear to be an ubiquitous finding across all RA gut microbiome studies, [16–21] including our own previous work [24]. Additionally, researchers have failed to replicate their findings of Prevotellaceae overabundance in larger cohorts [25]. Increasingly, numerous other bacteria have also been associated with RA development, including *Subdoligranulum* [26], *Lactobacillus*, [11,20] *Bacteroides* species [14,17] and *Streptococcaceae* [20,21], among others.

EULAR, the European Alliance of Associations for Rheumatology, defines at risk populations for RA based on a combination of genetic, environmental and immunological factors that predispose individuals to RA [27]. Therefore, heterogeneous RA at risk populations and confounding within microbiome studies further complicate the RA gut microbiome debate. This work examines a comprehensively phenotyped cohort of anticyclic citrullinated protein-positive (anti-CCP+) at risk individuals without clinical arthritis. We combined cross-sectional and longitudinal investigations of the gut microbiome with extensive clinical data, adjusting for potential confounding, to uncover microbial associations in the early RA continuum, with particular attention to Prevotellaceae strains.

METHODS

Patient and public involvement statement

Before starting this research, patients contributed to the study's design and feasibility, influencing recruitment, sample return and dissemination of outputs through a patient-focused discussion group at Chapel Allerton Hospital (CAH), Leeds. During the study, participant feedback on the stool collection kit was gathered and led to its redesign.

Study design and participants

This study used both cross-sectional and longitudinal approaches, involving participants from three groups: CCP+ at risk (n=124), newly diagnosed RA (NORA, n=7) and healthy controls (n=22). The study aimed to differentiate gut microbiome bacterial shifts due to confounding factors from those associated with RA progression, using comprehensive phenotypic data and serial stool samples as CCP+ at risk individuals progressed towards arthritis development.

The research was conducted at CAH, Leeds, UK. Individuals at risk of RA were selected from the Leeds CCP+ at risk cohort, as previously outlined [28]. These participants were recruited nationally from both primary and secondary care settings, presenting with recent musculoskeletal pain (onset <3 months) and positive for anti-CCP antibodies (CCP2 BioPlex 2200 Kit >2.99 IU/mL), but with no evidence of clinical synovitis/arthritis.

Within the CCP research clinic at CAH, individuals undergo longitudinal assessment for RA progression. An individual's risk of arthritis development was calculated using a validated RA risk severity score, based on symptoms, joint ultrasound, human leucocyte antigen (HLA) status and antibody titre [28], enabling comparison of gut microbiomes across varying RA risk.

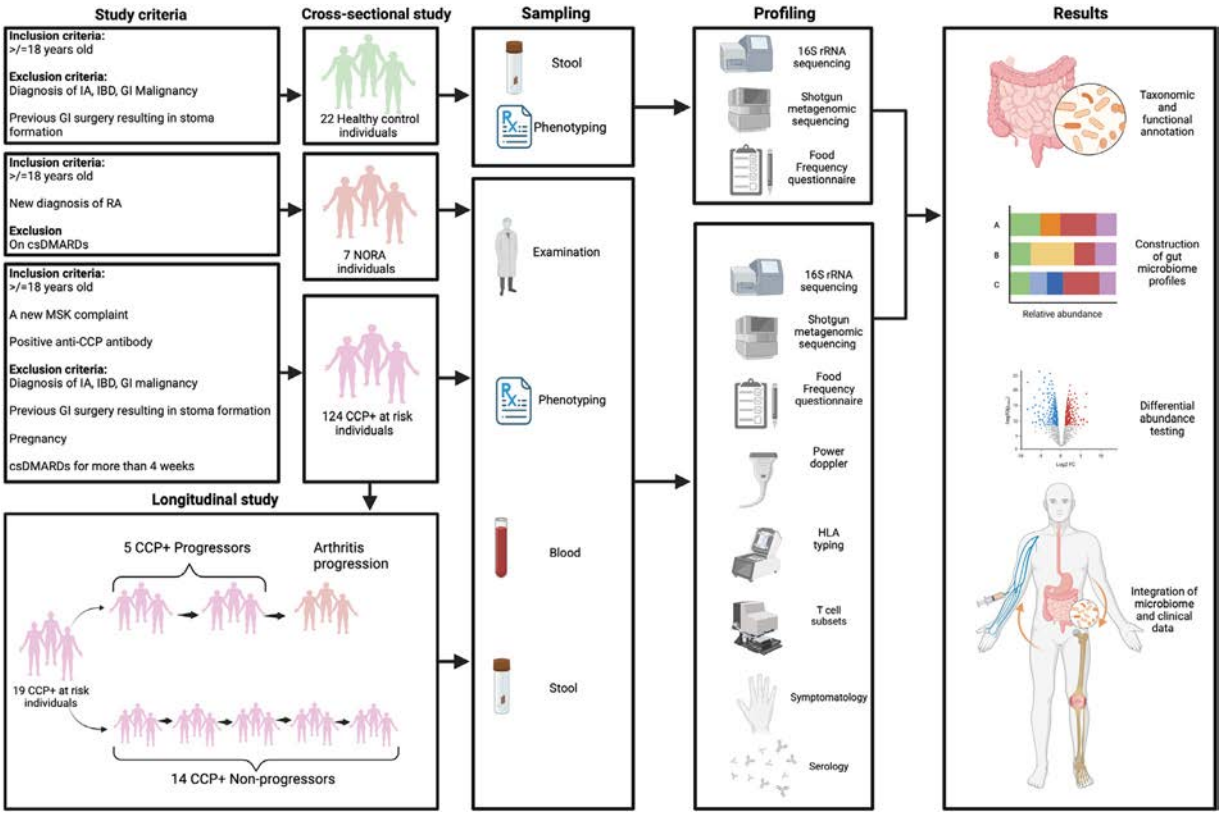


Figure 1. Study design and methodological flow chart. The figure depicts the recruitment of healthy controls, CCP+ at risk individuals and those with NORA, with corresponding criteria and subsequent sequential steps from sampling through molecular profiling and data analysis. Incorporates stool and blood analyses, sequencing methods and clinical assessments, culminating in the evaluation of microbial composition and its correlation with RA progression. CCP, cyclic citrullinated protein; csDMARDs, conventional synthetic disease-modifying antirheumatic drugs; GI, gastrointestinal; HLA, human leucocyte antigen; IA, inflammatory arthritis; IBD, inflammatory bowel disease; Log2 FC; Log2 fold change; MSK, musculoskeletal; NORA, new-onset rheumatoid arthritis; RA, rheumatoid arthritis.

Comprehensive data on demographics, medical history, medications, diet, bowel habit and lifestyle were also collected (figure 1).

As a comparison, individuals with NORA were recruited from general rheumatology clinics at CAH. Those diagnosed with RA and meeting the EULAR criteria, irrespective of CCP antibody status, were invited to participate in the cross-sectional study as participants with NORA. These individuals provided a single stool sample within 48 hours of their RA diagnosis, prior to starting DMARD therapy. Local healthy controls, free of musculoskeletal symptoms and representing diverse socioeconomic backgrounds, were also included, providing a single stool sample for the cross-sectional study. These controls were approximately matched with the CCP+ at risk cohort in terms of age, gender and smoking status.

Assessments

CCP+ at risk individuals underwent a baseline assessment, followed by quarterly follow-up appointments for the first year, with annual visits thereafter. The study endpoint was the development of clinical arthritis. A research helpline enabled prompt reporting of symptom changes for timely clinical evaluation. At baseline, demographic, medical, drug, social, stool and dietary histories [29] were collected. Clinical assessments at each visit included symptom review and joint assessment (78-joint count for tenderness and a 44-joint count for swelling).

Blood samples were collected at each visit for antibody testing, with the BioPlex 2200 machine used for anti-CCP2 testing and nephelometry for rheumatoid factor (RF). During the course of the study, a small number (n = 11) of low titre CCP+ individuals transitioned to a negative CCP+ titre at the time of gut microbiome sampling. High-resolution joint ultrasonography using power Doppler (PD) was performed at baseline, 6 monthly for the first year, then annually thereafter and intermittently on symptom exacerbation. PD assessment included scanning of 38 joints bilaterally (metacarpophalangeal, proximal interphalangeal and metatarsophalangeal joints). HLA-DR (Human Leukocyte Antigen-DR isotype) typing occurred during the baseline visit. T cell subpopulations were quantified using flow cytometry, adjusting counts for age. T cell counts were taken at baseline, 6 months and then annually.

Participants were given in-house stool collection kits for home sample collection, returning the kit at their subsequent clinic visit. Samples were kept at room temperature at home before being frozen at −80°C for long-term storage.

Molecular and bioinformatic methods

DNA was extracted from stool samples using the QIAamp DNA Stool Mini Kit (Qiagen, Germany). The University of Leeds' Next Generation Sequencing Facility at St James's University Hospital campus conducted sequencing for both 16S rRNA amplicon and shotgun metagenomic analysis. For 16S rRNA sequencing, the V4 region was amplified with specific primers and sequenced using Illumina MiSeq 2500, generating 2×250 base pair (bp) paired-end reads. Data were analysed using the DADA2 method [30] and SILVA reference database [31] within the QIIME2 [32] framework.

Shotgun metagenomic libraries were prepared with the Nextera XT DNA Library Prep Kit and sequenced on Illumina NextSeq 2000 using a 2×150 bp paired-end approach. The bioBakery workflow [33] was used for taxonomic and functional analysis of the shotgun sequencing data (see online supplemental materials, pages 1–3 for additional details).

All samples underwent 16S rRNA amplicon sequencing. In addition, 49 baseline samples, including CCP+ at risk individuals (n = 27), individuals with NORA (n = 7) and healthy controls (n = 15), also underwent shotgun sequencing. This dual approach was employed to validate the findings from the differential analysis of the 16S data, particularly to confirm the observed trends in *Prevotella* abundance and to explore any inferred functional differences arising from altered microbiome composition.

Analysis

Cross-sectional

In both sequencing methodologies, analyses were conducted in R (V.3.6.2) [34]. Diversity was calculated using phyloseq [35] and vegan [36] packages. For the 16S approach, a rooted phylogenetic tree was generated in QIIME2 [32], with alpha and beta diversity metrics calculated and Benjamini-Hochberg (BH)-adjusted p values (<0.05) determined via pairwise Wilcoxon rank-sum tests.

Permutational multivariate analysis of variance (MANOVA) was performed to assess the impact of metadata variables on beta diversity using three ecological metrics: Bray-Curtis, logged Bray-Curtis (Bray-Curtis was applied to scalar normalised counts (SNC) and logged data via the formula $\log_{10}(\mathrm{SNC} + 10^{-6}) + 6$) and unweighted UniFrac. Logged Bray-Curtis was used as an attempt to curtail outlying samples which might artificially amplify a signal within the Bray-Curtis index. Significant variables ($p<0.05$) were ranked by effect size and incorporated into a multivariate model to calculate cumulative variance explained.

For the shotgun method, a phyloseq object [35] was constructed using the enzyme commission number data set, and principal component analysis was performed to assess functional diversity of the microbiome. MaAsLin2 [37] was used for functional pathway analysis on prenormalised MetaCyc [38] annotated pathway abundance files, using an false discovery rate (FDR) of <0.05 as a significance threshold.

Differential bacterial abundance between clinically relevant groups was investigated across both sequencing methods using DESeq2, [39] adjusted for age and sequencing run. BLAST analyses supported further taxonomic annotation of significant Prevotellaceae strains. Differential abundance testing was performed at the genus level to determine higher taxonomic changes comparable to previous studies and at the strain level data to infer potential strain/subspecies information. In both methodologies, ggplot2 [40] facilitated data visualisation. Only taxa reaching a BH-adjusted p value of <0.05 were labelled. *Prevotella* strains of interest were identified based on *Prevotella* strains that reached a p value of <0.05 in the DESeq2 models from the cross-sectional study. The *Prevotella* strains of interest from the DESeq2 models and their associations were plotted using a chord plot.

Longitudinal

Longitudinal gut microbiome changes were analysed using 16S rRNA amplicon sequencing. Pairwise Bray-Curtis dissimilarities were constructed, and comparisons were limited to within-participant timepoints. These pairwise timepoints were categorised based on the time elapsed from RA diagnosis for progressors or from baseline for non-progressors. A logged Bray-Curtis distance matrix was used to construct a hierarchical tree using the 'dendrogram' and 'hclust' functions, applying the 'ward.D2' clustering algorithm. Mixed-effects linear models were created using the MaAsLin2 [37] package. Participants were assigned as the random-effect source in all models. To explore bacterial taxa, a centred log-ratio transformed data set was used at both the genus and strain levels.

RESULTS

The cross-sectional study encompassed 153 participants; their characteristics are detailed in table 1. In the RA at risk cohort, 70% were female, with a median anti-CCP antibody titre of 73 IU/mL. During the study period, 30 individuals progressed to RA, with a median interval of 9.6 months between stool collection and RA progression. The clinical attributes of progressors, non-progressors and NORA cohorts are outlined in table 2. Microbiome summary statistics are available in the online supplemental materials, page 4-5.

CCP+ at risk individuals are characterised by a low diversity state, which is associated with CCP antibody titre, RF positivity and HLA shared epitope status

Shannon diversity was notably reduced among CCP+ at risk individuals compared with healthy controls (Wilcoxon, p = 0.012; figure 2A). Diversity was reduced in both progressors and non-progressors, and appeared stratified by anti-CCP titre levels (figure 2A). Conversely, individuals with low or negative anti-CCP titres maintained diversity levels comparable to healthy controls. Both RF and HLA shared epitope (SE) positivity, known risk factors for arthritis development, were significantly linked to diminished diversity, as was steroid use (see online supplemental material, page 6). When examining the effects of these variables collectively in a multivariate model, none reached statistical significance (see online supplemental materials, page 6), suggesting that the individual contributions of these factors are less clear when considering the combined influence of all variables. Alpha diversity for measured lifestyle factors is presented in the online supplemental materials, page 6. Linear regression exploring variable effects on alpha diversity is presented in online supplemental materials, page 6-8.

Beta diversity analysis with respect to RA outcome status indicated no association between progression and enterotype (see online supplemental materials, page 9). Participants clustered according to dominant genera, primarily differentiating between *Prevotella*-dominated and *Bacteroides*-dominated microbiota. Progressors displayed both *Prevotella* and non-*Prevotella* predominant gut microbiome profiles. The analysis of 49 metadata variables using permutational analysis of variance (online supplemental materials, page 10-13) identified 15 variables with statistical significance ($p<0.05$) in at least one ecological metric (figure 2B), and subsequent multivariate models revealed sequencing run, age and vegetable intake as consistently significant factors across all three ecological distances, explaining microbial variance ranging from 8% to 10% (figure 2B). Alpha and beta diversity was calculated using 16S rRNA amplicon sequencing.

Progression from the at risk phase to arthritis development is associated with strain-specific changes in Prevotellaceae

A specific Prevotellaceae strain, ASV2058 (likely *P. copri*; BLAST analysis can be found in online supplemental materials,

Table 1
Characteristics of the cross-sectional study participants

Characteristics		CCP + at risk	NORA	Healthy
	Number of participants	124	7	22
Sociodemographic	Age (mean, SD)	53±12.5	57±15.6	54±15.9
	Female (%)	70 (88)	42 (3)	68 (15)
Diet	Food groups			
	Fruit servings/week	2	2	2
	Vegetable serving/week	2.1	1.8	1.8
	Oily fish*	1.9	2.6	1.8
	Fat†	2.397	2	2.235
	Non-milk extrinsic sugars†	2.321	2.6	2.412
	Dietary quality score (mean)	10.81	11	10.41
Lifestyle	Alcohol consumption (%)			
	None	32	0	22
	<14 units	44	14	40
	21 units	5.6	28	4.5
	>21 units	14.5	57	31
	Exercise type (%)			
	None	4	14	22
	Light	30	28	9
	Moderate	17	14	27
	Vigorous	6	14	18
	Exercise duration (average in minutes)	48	50	40
	Exercise frequency (weekly)	5	5	5
	Ever smoked (%)	41 (51)	42 (3)	36 (8)
Bowel habit	Stool type (average)	3	3	3
	Stool frequency per week (average)	8	8	7
Drugs	PPI (%)	31 (38)	42 (3)	28 (6)
	Regular laxatives (%)	0	0	0
	Antimicrobials within 3 months	0	0	0

CCP, cyclic citrullinated protein; NORA, new-onset rheumatoid arthritis; PPI, proton pump inhibitor.
* No intake = 0; 0–200 g/week = 1; ≥200 g/week = 2.
† ≥1½ × UK recommendations; 1–1½ × UK recommendations; ≤ UK recommendations.

Table 2
Baseline clinical characteristics of the cross-sectional musculoskeletal cohorts

Clinical characteristics		Progressors	Non-progressors	NORA
	Number of participants	30	94	7
MSK parameters	CCP level (U/mL) (median)	73	8.2	282
	CCP (mean, SD)	134±125	75±116	174±151
	CCP category (%)			
	Negative (<2.99 U/mL)	0	9 (9)	28 (2)
	Low (3–9 U/mL)	10 (3)	42 (40)	0
	High (9–90 U/mL)	46 (14)	23 (22)	14 (1)
	Very high (>90 U/mL)	43 (13)	24 (23)	57 (4)
	RF-positive (%)	70 (21)	30 (29)	100 (7)
	HLA-positive (%)	63 (19)	51 (48)	Unknown
	Tenderness (%)	50 (15)	40 (38)	100 (7)
	Early morning stiffness (%)	53 (16)	32 (31)	85 (6)
	Power Doppler (%)	43 (13)	14 (14)	Unknown
	Risk category (%)			
	Low	10 (3)	37 (35)	N/A
	Moderate	50 (15)	55 (52)	N/A
	High	40 (12)	7 (7)	N/A
	CRP category (% high)	20 (6)	10 (9)	57 (4)
	ESR category (% high)	53 (16)	35 (33)	42 (3)
	T-naïve cells category (% low)	53 (16)	26 (24)	Unknown
Medications	Polypharmacy	3.6	3.6	3.4
	NSAIDs (%)	60 (18)	50 (47)	71 (5)
	Steroid (%)	33 (10)	34 (31)	28 (2)
BMI (%)	Ideal	33 (10)	22 (21)	0
	Overweight	40 (12)	35 (33)	0
	Obese	23 (7)	29 (28)	100 (7)
Comorbidities (%)	Osteoarthritis	40 (12)	35 (33)	14 (1)
	Tendonitis	16 (5)	3 (3)	0
	Carpel tunnel syndrome	6 (2)	2 (2)	0
	Hypermobility	3 (1)	5 (5)	0
	Connective tissue disease	0	0	0
	Hypertension	20 (6)	12 (11)	14 (1)

(*continued*)

Table 2 (*Continued*)

Clinical characteristics	Progressors	Non-progressors	NORA
Ischaemic heart disease	6 (2)	3 (3)	0
Cerebrovascular disease	0	2 (2)	0
Diabetes	13 (4)	4 (4)	14 (1)
Renal disease	3 (1)	1 (1)	0
Chronic liver disease	0	1 (1)	0
Psoriasis	6 (2)	4 (4)	0
Chronic obstructive pulmonary disease	0	2 (2)	0

BMI, body mass index; CCP, cyclic citrullinated polypeptide; CRP, C reactive protein; ESR, erythrocyte sedimentation rate; HLA, human leucocyte antigen; MSK, musculoskeletal; N/A, not applicable; NORA, new-onset rheumatoid arthritis; NSAIDs, non-steroidal anti-inflammatory drugs; RF, rheumatoid factor.

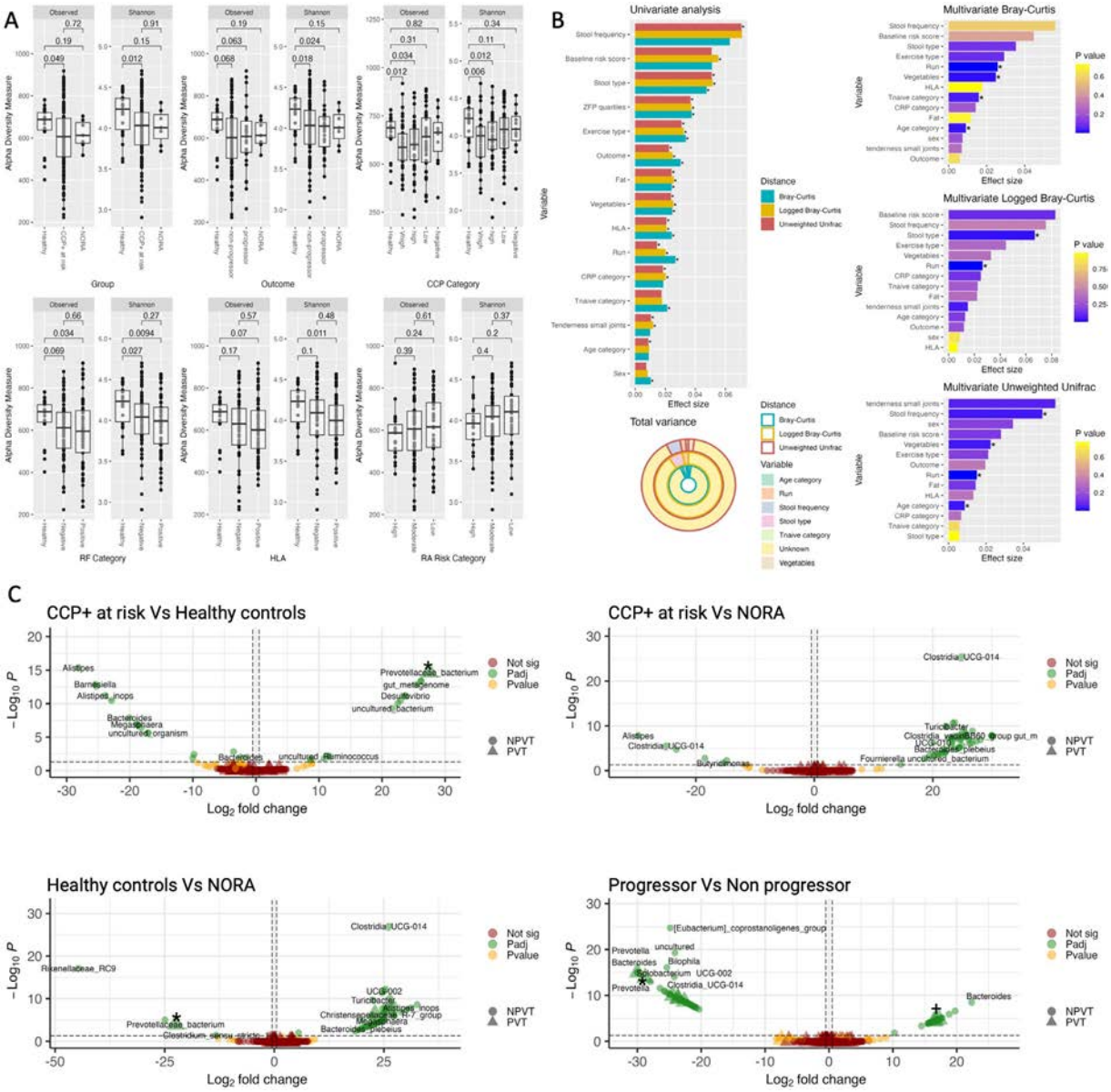


Figure 2. Gut microbiome diversity, variance and differential abundance from 16S rRNA amplicon sequencing. (A) Cross-sectional alpha diversity plots. Boxplots of alpha diversity as measured using observed and Shannon diversity metrics. BH-adjusted p values generated by pairwise Wilcoxon rank-sum test. (B) Microbial variance plots. Univariate analysis of significant ($p<0.05$) metadata variables using permutational ANOVA on three measures of beta diversity: Bray-Curtis, logged Bray-Curtis and unweighted UniFrac. Variables ordered by decreasing effect size. Total variance ring plot demonstrating the total known and unknown variance present within the gut microbiome. Individual multivariate permutational ANOVA, adjusted for significant variables identified in the univariate models, with significance reached in at least two ecological distances. Variables ordered by effect size. (C) Differentially abundant taxa between CCP+ at risk, healthy controls and NORA and between progressors and non-progressors. Pairwise volcano plots of plotting log2 fold change against −log10 p value, coloured according to significance. Vertical line denotes adjusted p value threshold; horizontal line denotes 0.5 log2 fold change. The shape denotes PVT or NPVT taxa. Taxonomic labels were generated to give the lowest taxonomic rank available in SILVA. *Denotes ASV2058. +Denotes ASV1867. (B) Outcome refers to progressor, non-progressor, NORA and healthy individuals. Run refers to sequencing run. ANOVA, analysis of variance; BH, Benjamini-Hochberg; CCP, cyclic citrullinated protein; CRP, C reactive protein; HLA, human leucocyte antigen; NORA, new-onset rheumatoid arthritis; NPVT, not Prevotellaceae; PVT, Prevotellaceae; RA, rheumatoid arthritis; RF, rheumatoid factor; ZFP, zonulin family peptide.

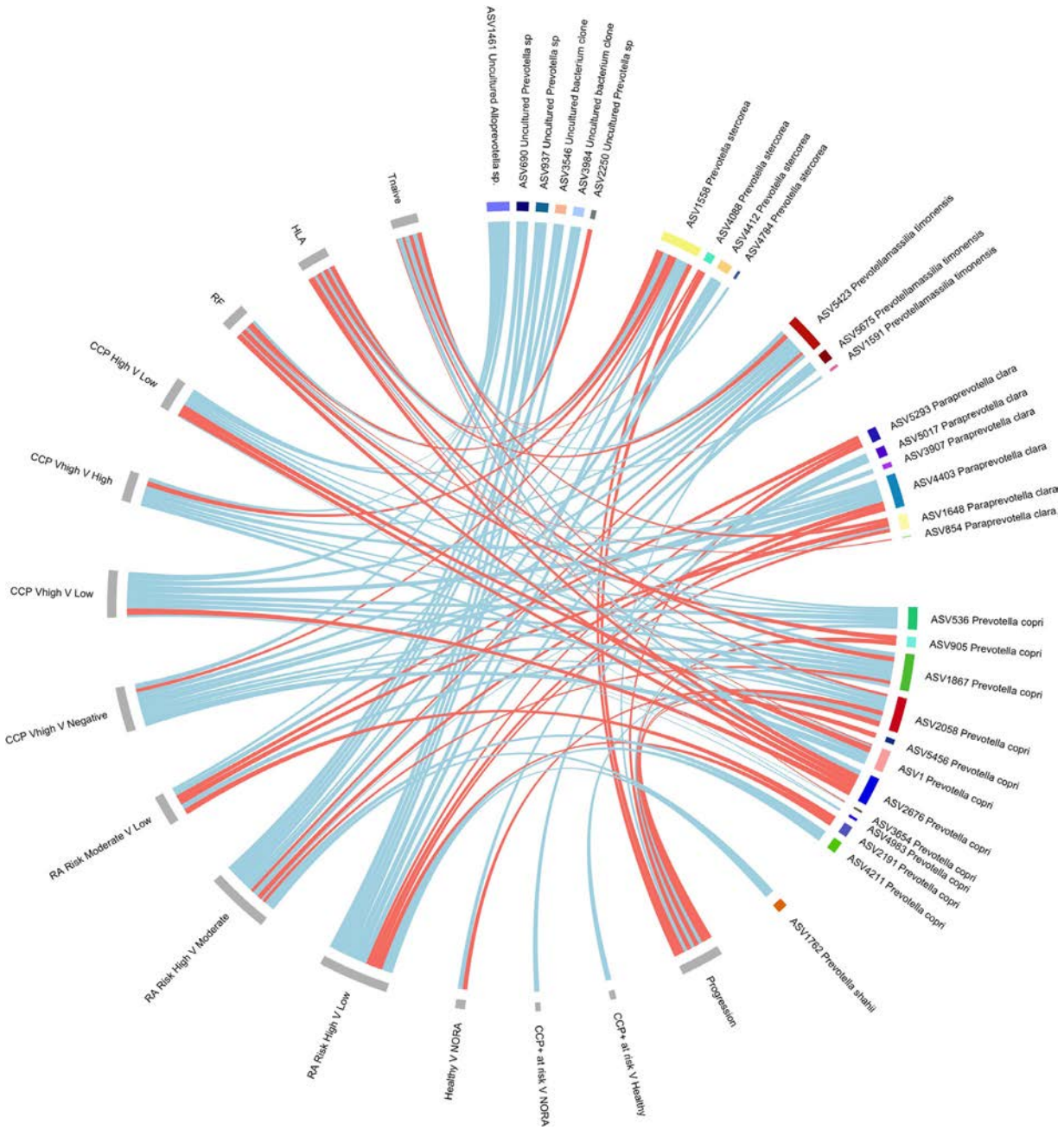


Figure 3. Prevotellaceae chord plot. Associations of Prevotellaceae strains reaching nominal significance from adjusted DESeq2 models from 16S rRNA amplicon sequencing. Metadata variables are depicted in grey and bacterial strains designated by colour. Chord links are coloured according to increased abundance (red) or decreased abundance (blue). The width of each chord is scaled according to the log2 fold changes in abundance. CCP, cyclic citrullinated protein; HLA, human leucocyte antigen; NORA, new-onset rheumatoid arthritis; RA, rheumatoid arthritis; RF, rheumatoid factor; V, versus.

page 30-31), was enriched in the at risk population compared with healthy controls. This same strain (ASV2058) was also found to be overabundant in individuals with NORA compared with healthy controls, with no significant difference in this putative *P. copri* strain between at risk and NORA (16S rRNA amplicon sequencing; see figure 2C).

Models were constructed to investigate a baseline RA progression signature within the CCP + at risk cohort by comparing progressors with non-progressors. Strain-level analysis demonstrates concordance with genus-level results (see online supplemental materials, page 14). Strain-specific phenomena are noted with both enrichment (three strains) and depletion (five strains) of Prevotellaceae-specific strains associated with progression (16S rRNA amplicon sequencing; see figure 2C). ASV2058 is included as one of those five depleted strains.

A phenomenon of strain-specific enrichment and concomitant depletion of Prevotellaceae strains was identified across multiple factors associated with higher risk of arthritis development, including anti-CCP antibody titre, RF positivity, HLA SE positivity, early immune dysregulation (T-naïve cells) and RA risk score (see online supplemental materials, page 15-19). All Prevotellaceae reaching nominal significance are detailed in online supplemental materials, page 20-29. Figure 3 provides a summary of Prevotellaceae strains found to be increased or decreased within the RA at risk population. Table 3 shows the proportion of each group containing Prevotellaceae and the proportion of Prevotellaceae strains of interest per group. From figure 3, Prevotellaceae strains ASV2058 and ASV1867 both had eight connections. RA risk category showed the highest number of associations with Prevotellaceae. Among all

Table 3
Proportion of Prevotellaceae presence

Group	Proportion containing any Prevotellaceae	Proportion containing Prevotellaceae strains of interest	χ^2 test
Progressor	0.38	0.87	2.32e-15
Non-progressor	0.75	0.67	0.0935
NORA	0.22	0.25	0.423
Healthy	0.09	0.35	0.03

NORA, new-onset rheumatoid arthritis.

Prevotellaceae, putative *P. copri* strains had the most associations with RA clinical variables; however, other strains were also implicated, such as *Alloprevotella*, *Paraprevotella clara*, *Prevotella stercorea*, *Prevotellamassilia timonensis* and *Prevotella shahii*.

Within the shotgun metagenomic data set, we assessed for the presence of the *P. copri* transposon described by Nii *et al* [41] but found no significant association with CCP positivity, disease progression or HLA status (online supplemental materials, page 50-51). However, we observed a higher number of transposon matches in individuals with specific *Prevotella* strains.

The shotgun sequencing data set comprising 27 CCP+ at risk (17 progressors), 7 NORA and 15 healthy controls corroborated an increase in Prevotellaceae species among at risk individuals relative to healthy controls (onlineonline supplemental materials, page 32-33) and strain-specific changes of Prevotellaceae associated with clinical variables (see online supplemental materials, page 35-39).

In CCP+ at risk individuals, gut microbiome alterations can begin 10 months before arthritis development

Due to the observed decrease in alpha diversity preceding RA onset (see online supplemental materials, page 44), pairwise dissimilarities between timepoints were assessed to examine gut microbiome stability using 16S rRNA amplicon sequencing. These dissimilarities were categorised into intervals related to arthritis development (ie, progression), either retrospective for progressors or prospective from baseline for non-progressors, and presented in figure 4A. Time to progression spanned 0−15 months preceding arthritis onset; the time for non-progressor from baseline to final follow-up was at least 12 months.

CCP+ progressors (ie, those who developed arthritis) exhibited the most significant instability ($p=0.044$, BH-adjusted t-test), with median Bray-Curtis dissimilarities exceeding 0.6 in the period immediately leading up to arthritis onset. This pattern indicates that gut microbiome structural changes commence around 10 months before clinical arthritis development and may extend beyond diagnosis. In contrast, the period from 10 to 15 months prior to RA onset and among non-progressors demonstrated relative stability, with lower Bray-Curtis dissimilarities suggesting more consistent microbiome structures over time and less variability between samples. Clinical, serological and radiological findings from the longitudinal participants are provided in the online supplemental materials, page 40-43.

Hierarchical clustering demonstrates modest microbial divergence prior to RA onset

To determine if samples at the progression timepoint were structurally more similar to each other or to their respective preceding samples (16S rRNA amplicon sequencing), a hierarchical tree was constructed. The resulting circular dendrogram showed higher structural similarity within individual successive samples than between different individuals. However, a notable divergence was observed between progression samples and their immediate predecessors, indicated by increased branch lengths in the dendrogram, such as seen with participant 8 and participant 3 (figure 4B). This was further explored using a multidimensional scaling analysis (online supplemental materials, page 44-45), which corroborates dispersion of the coordinates prior to RA diagnosis.

The gut microbiome of individuals progressing to arthritis accumulates pathobionts over time

Mixed-effects linear models on 16S rRNA amplicon sequencing data were used to analyse temporal bacterial fluctuations in the gut microbiome as some individuals progressed to arthritis. Each participant was defined as the random effect to accommodate correlations across multiple timepoints. The models analysed data at both the genus and strain levels (figure 4C).

At the genus-level analysis, six taxa were associated with samples collected at the time of RA diagnosis, termed progression (figure 4C), while three taxa were associated to samples preceding RA onset, termed progressors. Significantly, two taxa from the Lachnospiraceae family showed consistent depletion in both progression and progressors samples (figure 4C). Strain-level analysis corroborated these findings, demonstrating a decrease in ASV3244, a specific Lachnospiraceae strain, in both progressors and at progression. In contrast, certain bacterial strains were found to increase in the gut microbiome of RA progressors, especially at the progression point (figure 4C). Detailed outputs are enumerated in the online supplemental materials, page 46-49. Notably, ASV1026, identified as *Prevotella* and enriched in the progression samples, was identified as *P. copri* through BLAST analysis.

NORA gut microbiome is associated with enhanced amino acid and energy metabolism

MaAsLin linear modelling (FDR q value <0.25) using shotgun metagenomic sequencing identified differential abundance in 18 MetaCyc pathways when comparing at risk individuals (n=27), individuals with NORA (n=7) and healthy controls (n=15), with healthy controls serving as the reference group for all comparisons. Eleven pathways exhibited elevation in the NORA gut microbiome, while six showed reduction in healthy controls, as presented in figure 5A. There were no differentially significant pathways identified in CCP+ at risk group when compared with NORA or healthy controls. The MaAsLin analysis also identified key microbial features, including MetaCyc pathways, enzymes and gene families (figure 5B), that differ between individuals in various CCP categories (very high, high, low) when compared with healthy controls (online supplemental materials, pages 50-51). Notably, some enzymes, such as cobalt precorrin-5B (C.1) methyltransferase, were significantly associated with more than one CCP category (see figure 5B). Functional analysis of HLA, RF and RA risk category did not yield significant results.

DISCUSSION

This study presents a detailed examination of the gut microbiome's alterations before the onset of RA, revealing community-wide changes that extend beyond key organisms, both

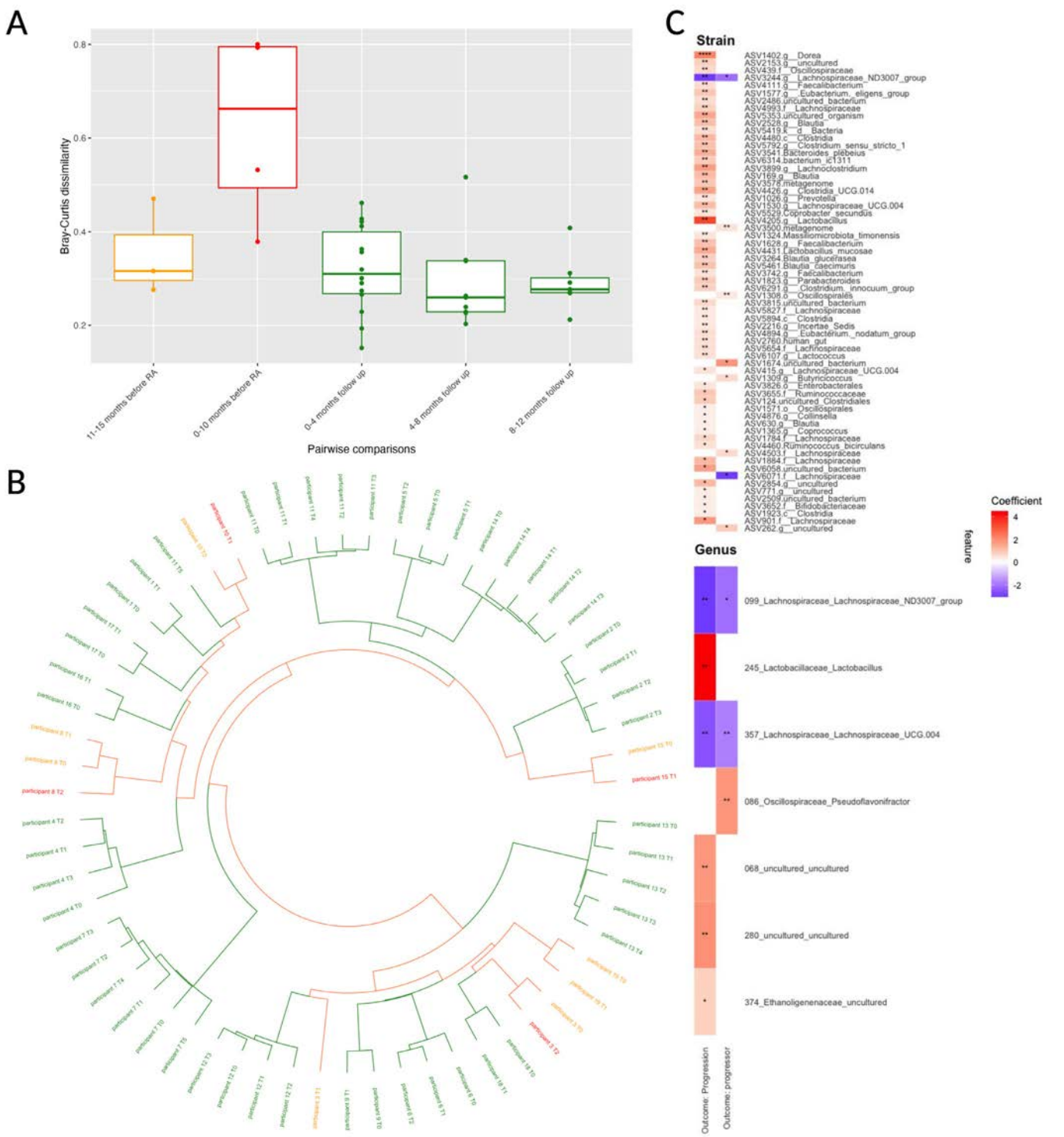


Figure 4. Temporal microbial analysis of CCP+ at risk individuals. (A) Pairwise Bray-Curtis dissimilarity boxplots from 16S rRNA amplicon sequencing. Boxplot of pairwise Bray-Curtis dissimilarities grouped according to time from progression: 0–10 months (n = 4) and 11–15 months (n = 3); or time from baseline for non-progressors: 0–4 months (n = 14), 4–8 months (n = 6) and 8–12 months (n = 6) of follow-up. (B) Circulated dendrogram of longitudinal samples. Constructed using logged Bray-Curtis dissimilarity index and clustered using Ward D2 methodology. Samples coloured according to outcome: non-progressors (green), progressors (orange) and progression (red). (C) Mixed-effects linear modelling with bacterial profiles. Linear model of top 50 strains with significant associations in heatmap, plotted with coloured according to coefficient; p values are indicated as follows: *p<0.25, **p<0.05, ***p<0.01, ****p<0.001. Progression denotes sampling at diagnosis of RA. Progressor denotes sampling prior to progression. All plots constructed using 16S RNA amplicon sequencing. CCP, cyclic citrullinated protein; RA, rheumatoid arthritis.

confirming and expanding on current literature. The work presented here represents, to the best of our knowledge, the largest, most homogenous RA at risk cohort, combining unique longitudinal and microbial variance analyses. Our study provides several novel insights, including decreased alpha diversity in the at risk phase, confirmation of *Prevotella* overabundance in at risk individuals and that *Prevotella* abundance appears to be associated with the underlying risk profile of the individual, which is consistent despite adjustment for common microbiome confounders. Finally, we propose a potential timeline for when gut microbiome changes occur as individuals develop RA.

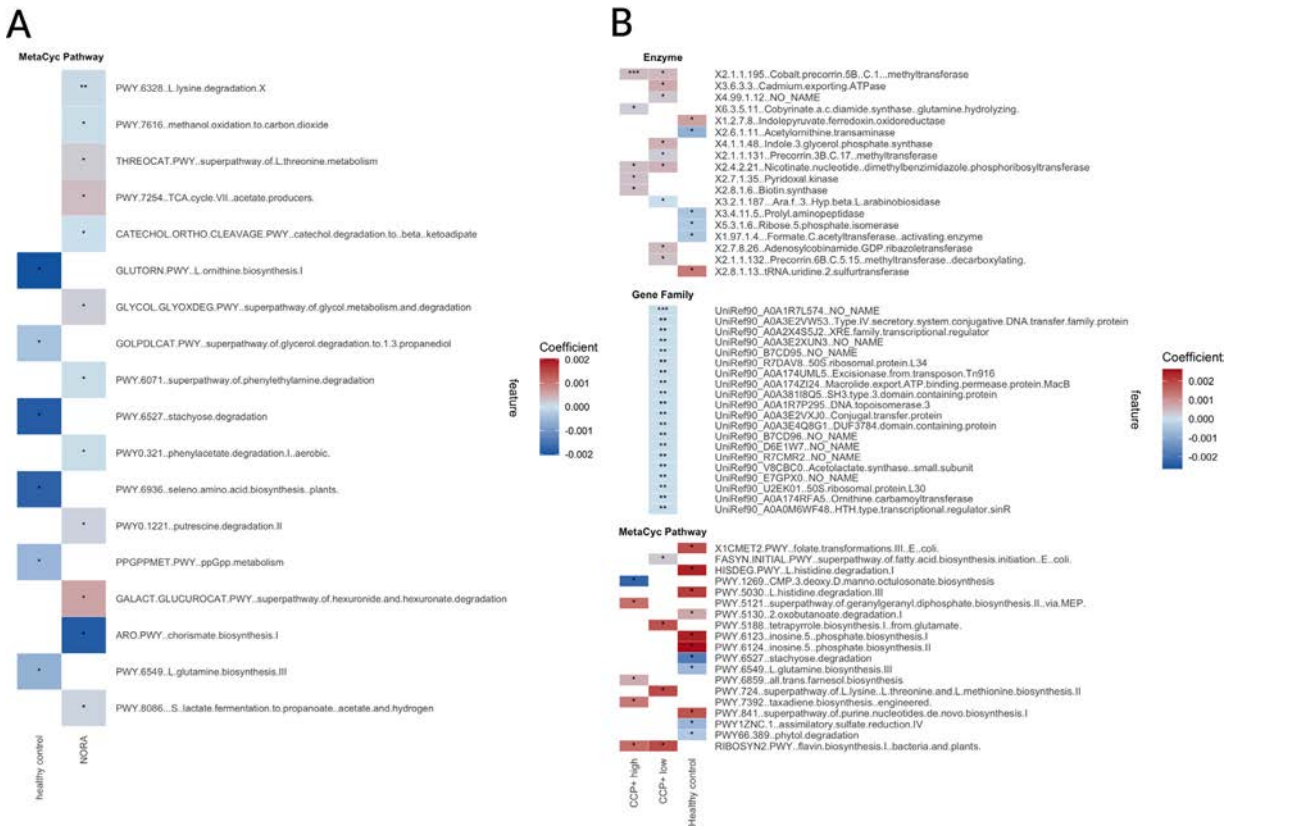


Figure 5. Functional analysis of enzymes, gene families and MetaCyc pathways from shotgun metagenomic sequencing. (A) Heatmap of MetaCyc pathways according to group. Heatmap represents MetaCyc pathways differentially abundant in NORA and healthy controls. (B) Heatmap of enzyme commission numbers, gene families and MetaCyc pathways of anti-CCP categories. Asterisks indicate the level of statistical significance: *p<0.25, **p<0.05, ***p<0.01. The colour gradient represents the coefficient values. Constructed using shotgun metagenomic data. CCP, cyclic citrullinated protein; NORA, new-onset rheumatoid arthritis.

Notably, we identified significantly decreased alpha diversity in CCP+ at risk individuals compared with healthy controls (figure 2A). This was previously unidentified in three similar studies [8,24,25]. These differences between our previous work and this study could be explained by considerable different cohort size (25 vs 124 CCP+ at risk, respectively). Additionally, our initial study relied on older operational taxonomic unit (OTU) clustering, which is less sensitive than newer denoising methodologies. Both Alpizar-Rodriguez *et al* [8] and Gilbert *et al* [25] have previously investigated the at risk phase of RA but there are considerable differences between the RA at risk populations included in these studies. RA risk can be identified through a combination of factors, including antibody status, symptoms and genetics. The cohort examined by Alpizar-Rodriguez *et al* [8] was characterised by positivity for anti-CCP and/or RF, along with the presence of symptoms. These symptoms were either assessed through a disease screening questionnaire or identified via a diagnosis of undifferentiated arthritis (UA). This approach results in a more diverse at risk population, potentially more advanced in the RA disease continuum due to the inclusion of individuals with UA. Furthermore, it also relied on older OTU analysis. Gilbert *et al* [25] used the same population; however, risk was subdivided into four categories, with varying risks including genetics, autoantibody presence and/or symptoms, again representing a heterogeneous population.

Interestingly, we showed alpha diversity was stratified according to anti-CCP antibody level (figure 2A), where individuals with high and very high anti-CCP levels had significantly lower alpha diversity, as defined by both observed and Shannon diversity index. In contrast, Gilbert *et al* [25] did not identify reduced alpha diversity in those at risk of RA. However, within the symptomatic and asymptomatic autoimmune groups, 55% and 70%, respectively, were anti-CCP antibody-negative, reflecting a very different population from our study. While eligibility for the Leeds at risk cohort has been defined through anti-CCP antibody positivity, longitudinal follow-up has revealed a small number of individuals (n=11) transition from low titre positive (anti-CCP2 >2.99 IU/mL and <9 IU/mL) to a negative anti-CCP antibody status. Low titre individuals and those converting to negative titre show preserved alpha diversity, similar to healthy controls (figure 2A), which corresponds with the findings of Gilbert *et al* [25]. Comparison of the NORA group with healthy controls did not show significantly decreased alpha diversity, and this may be reflective of the small numbers within this cohort and the heterogeneity in terms of antibody positivity (NORA 73% anti-CCP antibody-positive, 100% RF-positive). However, this result is in keeping with previous work, where others have failed to identify decreased alpha diversity, as measured by Shannon diversity index in NORA compared with healthy controls [9,11]. It is worth noting that we also observed decreased Shannon diversity with steroid use. However, given steroid use is pronounced in the symptomatic group (which tends to have higher CCP titres, HLA and RF positivity), it is impossible to know what portion of change is disease-related versus drug-related.

To understand the factors influencing microbial variability, a permutational MANOVA was conducted. This analysis identified age, vegetable intake and sequencing run as potential confounders (refer to figure 2B). Age stratification, ranging from 22 to 80 years in this cohort, aligns with existing literature [42–44]

as a critical factor in microbiome–disease correlations and was therefore incorporated into the statistical model. However, vegetable intake, while identified as a potential confounder, was not adjusted for in the model to avoid overcorrection that may obscure significant microbiome–disease associations. Considering the hypothesis that diet, including vegetable consumption, might influence the gut microbiome's role in RA-related autoimmunity, adjusting for vegetable intake could potentially conceal key bacterial associations that are either protective or harmful. Consequently, the study refrained from adjusting for vegetable intake in the analyses. The same rationale was applied to adjustment of intestinal barrier integrity (zonulin family peptide; see figure 2B). Given the relatively high frequencies of non-steroidal anti-inflammatory drugs, proton pump inhibitors and steroid use within the at risk phases of RA, we investigated their effect on the gut microbiome. None of these variables showed a significant effect in the permutational multivariate analysis, likely due to the pro re nata use of medications coded as 'ever used' rather than 'current use'.

The study highlights a distinct pattern of strain-specific enrichment and concurrent depletion of Prevotellaceae in the RA at risk phase, aligning with previous findings of Prevotellaceae overabundance in CCP+ at risk individuals compared with healthy controls (figure 2C). A specific strain of Prevotellaceae, ASV2058, showed significant enrichment in the CCP+ at risk group compared with healthy controls. BLAST analysis suggested this strain is likely *P. copri*. Notably, this enrichment was also observed in individuals with NORA compared with healthy controls. Intriguingly, ASV2058 was found to be less abundant in progressors versus non-progressors. Its abundance increased with higher CCP antibody titres and HLA positivity, but decreased in individuals with abnormal T-naïve cells. ASV1867, a second putative *P. copri* strain, was increased in progressors at baseline, positively correlated with CCP titres and abnormal T cells, but was negatively associated with HLA SE positivity. These findings might suggest that different strains of *P. copri* may have diverse roles in RA progression. One interpretation of these findings could be that ASV2058 is an additional risk factor for RA, associated with known serological and genetic risk markers, and in this context a lower threshold is required to induce early immune dysregulation and progression. Conversely, ASV1867 is not associated with genetic risk, and in this context relatively higher amounts are required for early immune dysregulation and RA progression.

Longitudinal analysis revealed that RA progressors tend to have a more unstable gut microbiome (figure 4A) compared with non-progressors. A median Bray-Curtis dissimilarity exceeding 0.6 was observed in progressors within 10 months prior to RA diagnosis (figure 4A), indicating a considerable shift in the gut microbiome during this period. When comparing the time periods of 0–10 months before RA and 11–15 months before RA, the gut microbiome appears relatively stable until 10 months before RA onset (figure 4A). The median Bray-Curtis index for the period of 11–15 months (0.3) aligns with values observed in non-progressors, suggesting significant changes may occur closer to RA onset. This increase in Bray-Curtis values during the period of 0–10 months may reflect the influence of RA development and associated inflammatory responses on the gut microbiome. However, for participant 3, both samples taken within 9 months before RA diagnosis showed a Bray-Curtis dissimilarity index of 0.8 (figure 4A), despite clinical reviews excluding RA progression at 9 and 6 months prior. This indicates that changes in the gut microbiome were occurring before the onset of clinical arthritis, although environmental influences cannot be ruled out. Similar patterns of beta diversity volatility have been noted in patients with inflammatory bowel disease around flare-ups, where increased instability correlated with heightened inflammation [45].

Our finding of increased putrescine degradation and threonine synthesis in the NORA cohort (figure 5A) is consistent with previous work [9,11]. Pathways such as L-lysine degradation, L-threonine metabolism and the tricarboxylic acid (TCA) cycle (acetate producers) were notably more active in the NORA group compared with healthy controls, suggesting an upregulation of amino acid metabolism and energy production in these individuals. Additionally, pathways involved in glycol metabolism and degradation, as well as phenylacetate degradation, were also elevated in individuals with NORA, indicating potential alterations in energy utilisation and aromatic compound processing. These findings suggest that individuals with NORA exhibit distinct metabolic shifts, particularly in pathways related to amino acid and energy metabolism, which could reflect underlying differences in metabolic demands or stress responses associated with inflammation.

Investigation of gut microbiome function in relation to anti-CCP antibody titre identified associations between several enzymes, gene families and metabolic pathways, aligning with literature showing that ACPA antibody titre affects microbiome function [46]. For instance, the increase of type IV secretory system conjugative DNA transfer family protein, XRE family transcriptional regulator, DNA topoisomerase and excisionase in CCP+ high and low individuals suggests a coordinated alteration in microbial gene expression. Additionally, enzymes such as cobalt precorrin-5B (C.1) methyltransferase, also elevated in both CCP+ high and low individuals, responsible for the biosynthesis in cobalamin (vitamin B_{12}), indicate potential adjustment in microbial vitamin synthesis. The activity of other enzymes, such as cadmium-exporting ATPase and pyridoxal kinase, points to altered metal ion homeostasis and coenzyme metabolism. Histidine degradation, inosine 5'-phosphate biosynthesis and folate transformations reveal alterations in amino acid metabolism, nucleotide synthesis and one carbon metabolism (oxidative stress), reflective of our findings in the NORA group. It is worth noting, although comparable dietary quality scores suggest that there were no major deviations in diet quality, including protein intake. We do recognise that direct quantification of protein consumption could provide more precise insights into its potential impact on amino acid metabolism.

We specifically assessed for the presence of the transposon described by Nii *et al* [41] in our shotgun metagenomics data. Our results did not show any significant relationship between the presence of this transposon and clinical factors such as CCP positivity, disease progression or HLA status (online supplemental materials, page 51-52). The lack of association with these variables suggests that while *Prevotella* strains with this transposon are indeed present in the cohort (including healthy controls) and potentially relevant, their presence alone does not correlate with these specific clinical outcomes in our cohort. We did identify a higher number of transposon matches in individuals who possessed our identified *Prevotella* strains of interest. This suggests an interesting association between the presence of specific Prevotellaceae strains which are associated with RA clinical variables and progression and an increased number of matches to a *P. copri* transposon sequence, and raises the question of whether this is merely a reflection of overall *Prevotella* abundance or if it points to a more specific biological relationship; this warrants further work beyond the scope of this paper.

The limitations of this research, particularly the small longitudinal sample size and the lack of 1:1 longitudinal comparison

between CCP+ at risk and healthy controls, temper the generalisability of the results and increase the chance of spurious associations. This should be considered when interpreting the findings and calls for further research with larger longitudinal cohorts. The heterogeneity in the NORA cohort reflects the practical constraints of recruitment from standard of care clinics, which may limit the interpretability of direct comparisons. Moreover, extending the analysis beyond the bacterial microbiome to include viromes, mycobiomes and archaeomes could unveil additional facets of microbial influence on RA pathogenesis. The absence of integrated transcriptomic or metabolomic data restricts our interpretation to potential rather than confirmed metabolic activity.

In summary, we have shown that individuals at risk of RA harbour a distinctive gut microbial composition, including but not limited to an overabundance of Prevotellaceae species. This microbial signature is consistent and correlates with traditional RA risk factors. Longitudinal examination shows a dynamic microbial environment preceding RA onset. Further research into this late phase of disease development is merited, especially given the potential of the gut microbiome as a target for prevention, including in high-risk individuals with imminent arthritis.

Acknowledgments

We extend our gratitude to the participants of the Leeds at risk cohort and the patient discussion groups. We also wish to acknowledge the work of Professor F Ponchel with T cell subset analysis and Professor A Kastbom with ZFP quantification. Figure 1 was created with BioRender.com.

Contributors

The study was designed by PE, KM, MHW and CMR. Sample and data collection was conducted by KM and CMR. Sample processing was completed by CMR. Data analysis was completed by CMR and supervised by IBJ. All authors contributed to writing the manuscript and approved it for publication. CMR is the guarantor.

Funding

CMR was funded to complete this work by a personal fellowship from Versus Arthritis (22294) and Leeds Cares (9R01-A1210) and is currently funded by a National Institute for Health Research (NIHR) Clinical Lectureship. The Leeds CCP clinic is supported by the NIHR Leeds Biomedical Research Unit.

Competing interests

CMR has received funding from Versus Arthritis and Leeds Cares and in-kind support from 4D Pharma PLC. KM has received grants from Gilead, Lilly, Serac Healthcare, AstraZeneca and DeepCure; consulting fees from AbbVie, UCB, Lilly, Galapagos, Serac Healthcare, Zura Bio and DeepCure; and honoraria from AbbVie, UCB, Lilly, Galapagos, Serac Healthcare, Zura Bio and DeepCure. PE has received grants from AbbVie, BMS, Lilly, Novartis, Pfizer and Samsung; and consulting fees from AbbVie, Activia, AstraZeneca, BMS, Boehringer Ingelheim, Galapagos, Gilead, Immunovant, Janssen, Lilly and Novartis.

Patient and public involvement

Before starting this research, patients contributed to the study's design and feasibility, influencing recruitment, sample return and dissemination of outputs through a patient-focused discussion group at Chapel Allerton Hospital, Leeds. During the study, participant feedback on the stool collection kit was gathered and led to its redesign.

Patient consent for publication

Not required.

Ethics approval

This study involves human participants and was conducted at Chapel Allerton Hospital (CAH), Leeds, UK, with ethical approval granted by the Leeds, West Yorkshire Research Ethics Committee (reference: 06/Q1205/169). Participants gave informed consent to participate in the study before taking part.

Data availability statement

Raw sequencing data available upon reasonable request from the corresponding author.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1136/ard-2024-226362.

Orcid

Christopher M. Rooney: http://orcid.org/0000-0002-3640-853X
Kulveer Mankia: http://orcid.org/0000-0002-7945-6582
Paul Emery: http://orcid.org/0000-0002-7429-8482

REFERENCES

[1] Holers VM, Demoruelle MK, Kuhn KA, et al. Rheumatoid arthritis and the mucosal origins hypothesis: protection turns to destruction. Nat Rev Rheumatol 2018;14:542–57. doi: 10.1038/s41584-018-0070-0.
[2] Sugiyama D, Nishimura K, Tamaki K, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis 2010;69:70–81. doi: 10.1136/ard.2008.096487.
[3] Mankia K, Cheng Z, Do T, et al. Prevalence of Periodontal Disease and Periodontopathic Bacteria in Anti-Cyclic Citrullinated Protein Antibody-Positive At-Risk Adults Without Arthritis. JAMA Netw Open 2019;2:e195394. doi: 10.1001/jamanetworkopen.2019.5394.
[4] Sokolova MV, Hartmann F, Sieghart D, et al. Antibodies against citrullinated proteins of IgA isotype are associated with progression to rheumatoid arthritis in individuals at-risk. RMD Open 2023;9:e002705. doi: 10.1136/rmdopen-2022-002705.
[5] Kastbom A, Roos Ljungberg K, Ziegelasch M, et al. Changes in anti-citrullinated protein antibody isotype levels in relation to disease activity and response to treatment in early rheumatoid arthritis. Clin Exp Immunol 2018;194:391–9. doi: 10.1111/cei.13206.
[6] van der Woude D, Rantapää-Dahlqvist S, Ioan-Facsinay A, et al. Epitope spreading of the anti-citrullinated protein antibody response occurs before disease onset and is associated with the disease course of early arthritis. Ann Rheum Dis 2010;69:1554–61. doi: 10.1136/ard.2009.124537.
[7] Schellekens GA, Visser H, De Jong BAW, et al. The diagnostic properties of rheumatoid arthritis antibodies recognizing a cyclic citrullinated peptide. Arthritis & Rheumatism 2000;43:155–63 http://dx.doi.org/10.1002/1529-0131(200001)43:1<155::AID-ANR20>3.0.CO;2-3.
[8] Alpizar-Rodriguez D, Lesker TR, Gronow A, et al. Prevotella copri in individuals at risk for rheumatoid arthritis. Ann Rheum Dis 2019;78:590–3. doi: 10.1136/annrheumdis-2018-214514.

[9] Scher JU, Sczesnak A, Longman RS, et al. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife 2013;2: e01202. doi: 10.7554/eLife.01202.
[10] Wells PM, Adebayo AS, Bowyer RCE, et al. Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. Lancet Rheumatol 2020;2:e418–27. doi: 10.1016/S2665-9913(20)30064-3.
[11] Zhang X, Zhang D, Jia H, et al. The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. N Med 2015;21:895–905.
[12] Round JL, Lee SM, Li J, et al. The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. Science 2011;332:974–7. doi: 10.1126/science.1206095.
[13] Rodrigues GSP, Cayres LCF, Gonçalves FP, et al. Detection of Increased Relative Expression Units of Bacteroides and Prevotella, and Decreased Clostridium leptum in Stool Samples from Brazilian Rheumatoid Arthritis Patients: A Pilot Study. Microorganisms 2019;7:10. 413 10.3390/microorganisms7100413.
[14] Jeong Y, Kim J-W, You HJ, et al. Gut Microbial Composition and Function Are Altered in Patients with Early Rheumatoid Arthritis. J Clin Med 2019;8:693. doi: 10.3390/jcm8050693.
[15] Maeda Y, Kumanogoh A, Takeda K. Altered composition of gut microbiota in rheumatoid arthritis patients. Nihon Rinsho Meneki Gakkai Kaishi 2016;39:59–63. doi: 10.2177/jsci.39.59.
[16] Breban M, Tap J, Leboime A, et al. Faecal microbiota study reveals specific dysbiosis in spondyloarthritis. Ann Rheum Dis 2017;76:1614–22. doi: 10.1136/annrheumdis-2016-211064.
[17] Vaahtovuo J, Munukka E, Korkeamäki M, et al. Fecal microbiota in early rheumatoid arthritis. J Rheumatol 2008;35:1500–5.
[18] Muñiz Pedrogo DA, Chen J, Hillmann B, et al. An Increased Abundance of Clostridiaceae Characterizes Arthritis in Inflammatory Bowel Disease and Rheumatoid Arthritis: A Cross-sectional Study. Inflamm Bowel Dis 2019;25:902–13. doi: 10.1093/ibd/izy318.
[19] Chiang H-I, Li J-R, Liu C-C, et al. An Association of Gut Microbiota with Different Phenotypes in Chinese Patients with Rheumatoid Arthritis. J Clin Med 2019;8:11. 1770 10.3390/jcm8111770.
[20] Picchianti-Diamanti A, Panebianco C, Salemi S, et al. Analysis of Gut Microbiota in Rheumatoid Arthritis Patients: Disease-Related Dysbiosis and Modifications Induced by Etanercept. Int J Mol Sci 2018;19:2938. doi: 10.3390/ijms19102938.
[21] Forbes JD, Chen C-Y, Knox NC, et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases-does a common dysbiosis exist? Microbiome 2018;6:221. doi: 10.1186/s40168-018-0603-4.
[22] Danckert NP, Freidin MB, Smith IG, et al. Treatment response in rheumatoid arthritis is predicted by the microbiome: a large observational study in UK DMARD-naïve patients. Rheumatology (Oxford) 2024:keae045. doi: 10.1093/rheumatology/keae045.
[23] Maeda Y, Kurakawa T, Umemoto E, et al. Dysbiosis Contributes to Arthritis Development via Activation of Autoreactive T Cells in the Intestine. Arthritis & Rheumatology 2016;68:2646–61. doi: 10.1002/art.39783.
[24] Rooney CM, Mankia K, Mitra S, et al. Perturbations of the gut microbiome in anti-CCP positive individuals at risk of developing rheumatoid arthritis. Rheumatology (Sunnyvale) 2021;60:3380–7. doi: 10.1093/rheumatology/keaa792.
[25] Gilbert BTP, Tadeo RYT, Lamacchia C, et al. Gut microbiome and intestinal inflammation in preclinical stages of rheumatoid arthritis. RMD Open 2024;10:e003589. doi: 10.1136/rmdopen-2023-003589.
[26] Chriswell ME, Lefferts AR, Clay MR, et al. Clonal IgA and IgG autoantibodies from individuals at risk for rheumatoid arthritis identify an arthritogenic strain of Subdoligranulum. Sci Transl Med 2022;14:eabn5166. doi: 10.1126/scitranslmed.abn5166.
[27] Gerlag DM, Raza K, van Baarsen LGM, et al. EULAR recommendations for terminology and research in individuals at risk of rheumatoid arthritis: report from the Study Group for Risk Factors for Rheumatoid Arthritis. Ann Rheum Dis 2012;71:638–41. doi: 10.1136/annrheumdis-2011-200990.
[28] Rakieh C, Nam JL, Hunt L, et al. Predicting the development of clinical arthritis in anti-CCP positive individuals with non-specific musculoskeletal symptoms: a prospective observational cohort study. Ann Rheum Dis 2015;74:1659–66. doi: 10.1136/annrheumdis-2014-205227.
[29] Cleghorn CL, Harrison RA, Ransley JK, et al. Can a dietary quality score derived from a short-form FFQ assess dietary quality in UK adult population surveys? Public Health Nutr 2016;19:2915–23. doi: 10.1017/S1368980016001099.
[30] Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 2016;13:581–3. doi: 10.1038/nmeth.3869.
[31] Pruesse E, Quast C, Knittel K, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 2007;35:7188–96. doi: 10.1093/nar/gkm864.
[32] Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852–7. doi: 10.1038/s41587-019-0209-9.
[33] Beghini F, McIver LJ, Blanco-Míguez A, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife 2021;10:e65088. doi: 10.7554/eLife.65088.
[34] Team RC. R: A language and environment for statistical computing. 2013.
[35] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8:e61217. doi: 10.1371/journal.pone.0061217.
[36] Dixon P. VEGAN, a package of R functions for community ecology. J Vegetation Science 2003;14:927–30. doi: 10.1111/j.1654-1103.2003.tb02228.x.
[37] Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. PLoS Comput Biol 2021;17: e1009442. doi: 10.1371/journal.pcbi.1009442.
[38] Caspi R, Altman T, Billington R, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res 2014;42:D459–71. doi: 10.1093/nar/gkt1103.
[39] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550. 550 10.1186/s13059-014-0550-8.
[40] Wickham H. Wiley interdisciplinary reviews: computational statistics 2011;3:180–5. doi: 10.1002/wics.147.
[41] Nii T, Maeda Y, Motooka D, et al. Genomic repertoires linked with pathogenic potency of arthritogenic Prevotella copri isolated from the gut of patients with rheumatoid arthritis. Ann Rheum Dis 2023;82:621–9. doi: 10.1136/ard-2022-222881.
[42] Biagi E, Nylund L, Candela M, et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. PLoS One 2010;5: e10667. doi: 10.1371/journal.pone.0010667.
[43] Jeffery IB, Lynch DB, O'Toole PW. Composition and temporal stability of the gut microbiota in older persons. ISME J 2016;10:170–82. doi: 10.1038/ismej.2015.88.
[44] Ghosh TS, Das M, Jeffery IB, et al. Adjusting for age improves identification of gut microbiome alterations in multiple diseases. Elife 2020;9:e50240. doi: 10.7554/eLife.50240.
[45] Halfvarson J, Brislawn CJ, Lamendella R, et al. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol 2017;2:17004. doi: 10.1038/nmicrobiol.2017.4.
[46] He J, Chu Y, Li J, et al. Intestinal butyrate-metabolizing species contribute to autoantibody production and bone erosion in rheumatoid arthritis. Sci Adv 2022;8:eabm1511. doi: 10.1126/sciadv.abm1511.

Ann Rheum Dis 84 (2025) 575–583

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Psoriatic arthritis

An interpretable machine learning approach for detecting psoriatic arthritis in a UK primary care psoriasis cohort using electronic health records from the Clinical Practice Research Datalink

Alexander Rudge[1], Neil McHugh[2,3], William Tillett[2,3], Theresa Smith[1,*]

[1] *Department of Mathematical Sciences, University of Bath, Bath, UK*
[2] *Department of Life Sciences, University of Bath, Bath, UK*
[3] *Royal National Hospital for Rheumatic Diseases, Royal United Hospitals, Bath, UK*

ARTICLE INFO

Article history:
Received 11 November 2024
Received in revised form 28 January 2025
Accepted 31 January 2025

ABSTRACT

Objectives: Develop an interpretable machine learning model to detect patients with newly diagnosed psoriatic arthritis (PsA) in a cohort of psoriasis patients and identify important clinical indicators of PsA in primary care.

Methods: We developed models using UK primary care electronic health records from the Clinical Practice Research Datalink (CPRD). The study population consisted of a cohort of (PsA free) patients with incident psoriasis who were followed prospectively. We used Bayesian networks (BNs) to identify patients who developed PsA using primary care variables measured prior to diagnosis and compared the results to a random forest (RF). Variables included patient demographics, musculoskeletal symptoms, blood tests, and prescriptions. The importance of each variable used in the models was evaluated using permutation variable importance. Model discrimination was measured using the area under the receiver operating characteristic curve (AUC) and the area under the precision-recall curve (PRAUC).

Results: We identified a cohort of 122,330 patients with an incident psoriasis diagnosis between 1998 and 2019 in the CPRD, of whom 2460 patients went on to develop PsA. Our best BN achieved an AUC of 0.823, and PRAUC of 0.221, compared to the AUC of 0.851 and PRAUC of 0.261 of the RF. Psoriasis duration, nonsteroidal anti-inflammatory drug prescriptions, nonspecific arthritis, nonspecific arthralgia, and C-reactive protein blood tests were all important variables in our models.

Conclusions: We were able to identify psoriasis patients at higher risk, and important indicators, of PsA in UK primary care. Further work is required to evaluate our model's usefulness in assisting PsA screening.

*Correspondence to Dr. Theresa Smith, Department of Mathematical Sciences, University of Bath, Bath, UK.
E-mail address: t.r.smith@bath.ac.uk (T. Smith).
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.01.051

WHAT IS ALREADY KNOWN ON THIS TOPIC

- The prodromal phase of psoriatic arthritis presents an opportunity for early detection.
- Bayesian networks (BNs) are an interpretable machine learning model well suited for clinical predictions.

WHAT THIS STUDY ADDS

- An interpretable BN to identify psoriasis patients at higher risk of psoriatic arthritis (PsA) using primary care data recorded prior to diagnosis.
- Highlights indicators of PsA in primary care electronic health records, including nonsteroidal anti-inflammatory drug prescriptions, and C-reactive protein (CRP) blood tests.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- Models developed using electronic health records could provide useful tools to assist with PsA screening in primary care.
- Future research on developing screening tools for identifying individuals at higher risk of PsA should include indicators for nonsteroidal anti-inflammatory drug prescriptions and CRP blood tests.

INTRODUCTION

Psoriatic arthritis (PsA) is an inflammatory arthritis affecting up to 1 in 5 people with psoriasis [1]. Psoriasis often precedes PsA [2], meaning that patients with psoriasis form an at-risk population for screening. PsA is an insidious disease, characterised by a period of nonspecific musculoskeletal symptoms [3], which can make the exact time of disease onset unclear. Recent efforts have focused on characterising the prodromal phase of PsA where symptoms may develop, representing an opportunity for early detection [4,5]. The heterogeneous nature of PsA can be difficult to distinguish from other musculoskeletal diseases such as osteoarthritis [6,7]. These factors, combined with the lack of a PsA-specific biomarker, can make PsA difficult to diagnose. Furthermore, it is known that delays to diagnosis are often associated with worsening disease outcomes [8,9]. Because of this, recent PsA research has included the development of clinical prediction tools that can recognise high-risk individuals. These prediction models aim to improve patient outcomes by reducing the time to diagnosis and allowing earlier interventions.

Eder et al [10] developed the Psoriatic Arthritis Risk Estimation Tool (PRESTO), a model for predicting 1- and 5-year PsA risk in patients with psoriasis, and, similarly, Ogdie et al [11] studied the prediction of 2-year risk. Both studies used secondary care cohorts of patients with psoriasis to fit binary regression models and may have been limited by a small sample size. Electronic health records (EHRs) can overcome this limitation, allowing more complex models to be developed. For example, Shapiro et al [12] used EHRs to develop a machine learning model based on gradient boosting trees to predict the development of PsA in both a psoriasis cohort and a general population cohort. Green et al [13] used EHRs to identify PsA in patients with psoriasis using a Bayesian network (BN), a probabilistic graphical model that is being increasingly used for clinical prediction tasks [14]. BNs are particularly appealing in the clinical domain as they can naturally discover interactions between variables in the data and visualise them in an interpretable manner.

In this study, we built upon the methodology presented by Green et al [13] to develop a BNs to identify psoriasis patients who developed PsA using variables measured prior to diagnosis. We sought to improve upon Green et al [13] by using an extended study period, resulting in a larger sample size, and including more clinical features such as blood tests and prescription data. We also explored a greater number of BNs classifiers and compared their performance to a random forest (RF). Our objective is to develop a model that could be used to identify patients with undiagnosed PsA in primary care and identify important primary care clinical observations associated with a diagnosis of PsA.

METHODS

Study design

We performed a UK-based cohort study of patients with incident psoriasis to investigate the factors that contributed to a diagnosis of PsA. Patients were followed from their date of psoriasis diagnosis until the earliest occurrence of the following: they developed PsA, the end of the study period, they turned 90 years of age, or they ceased contributing data.

Data source

We used data extracted from the Clinical Practice Research Datalink (CPRD) GOLD, a curated database of UK primary care EHRs. Routinely collected patient data are provided to the CPRD by participating general practices (GPs) and includes patient demographics, symptoms, diagnoses, prescriptions, vaccination history, laboratory tests, and referrals to hospital and secondary care. The population within the CPRD is generally representative of the UK population in terms of age, sex, and ethnicity [15], and data validation checks by the CPRD ensure that the data are up-to-standard (UTS). Medical observations are recorded by GPs using a coded thesaurus of clinical terms called Read codes, and drug prescriptions are recorded using the Gemscript product code system. Code lists used in this study were developed in consultation with rheumatologists (WT and NJM).

Study population

The study population consisted of a cohort of patients with incident psoriasis, identified in the CPRD between January 1, 1998, and December 31, 2019. Eligible participants were aged between 16 and 89 years (inclusive) and were permanently registered and contributing UTS data to the CPRD. Patients were excluded from the study if they had existing inflammatory arthritis, including PsA, rheumatoid arthritis, or ankylosing spondylitis, prior to their psoriasis diagnosis.

The date of psoriasis diagnosis was taken to be the date of the first code for psoriasis in a patient's medical history unless there was evidence of treatment for psoriasis before this date. If there was, then the diagnosis date was backdated to the first occurrence of any of the following: a phototherapy treatment, a prescription for a vitamin D analogue (except if a patient had a record of vitiligo), or a prescription for dithranol. A diagnosis was treated as incident if there was at least 1 year of data collection prior to the date of diagnosis, which has been shown to help differentiate between incident and historic diagnoses [16].

The date of PsA diagnosis was initially taken to be the date of the first code for PsA in a patient's medical record. For some patients, the diagnosis was backdated to the date of the first prescription for the conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) sulfasalazine, leflunomide, or

methotrexate. Methotrexate can also be used to treat severe cases of psoriasis; therefore, we did not backdate patients via methotrexate if there was evidence to suggest that it was prescribed by a dermatologist. We leveraged referral data to indicate the specialty from which a prescription originated, as this information is not recorded within the CPRD. We identified a patient's earliest methotrexate prescription that was within 180 days of a dermatology or rheumatology referral and excluded patients from backdating if the referral was from dermatology, with no rheumatology referral in the same period. Otherwise, patients were backdated using their earliest methotrexate prescription, even if no referral data were present. We found that referral data in the CPRD is often missing, and it was common to find patients with prescriptions for drugs such as methotrexate that are very likely to have been prescribed in secondary care, with no relevant referral information.

As PsA can present similarly to other musculoskeletal diseases, a patient with a single PsA code and no evidence of treatment for PsA was considered a misdiagnosis if the patient had 3 or more subsequent codes for either osteoarthritis or fibromyalgia.

Study variables

We identified variables recorded during the study period for use in the classification models. These included demographics of age, biological sex, body mass index (BMI), smoking, and alcohol status; musculoskeletal symptoms; prescriptions for nonsteroidal anti-inflammatory drugs (NSAIDs) and corticosteroids; blood tests measuring C-reactive protein (CRP) levels and plasma viscosity (PV); psoriasis treatments made up of phototherapy treatment and drug prescriptions; and psoriasis duration. Any continuous variables were discretised, including a category for missing values. Additional details on how these variables were defined is provided in the Supplementary Material. The musculoskeletal symptom groups used in the analysis are listed in Supplementary Table S1, and the NSAIDs and corticosteroid drugs are listed in Supplementary Table S2.

Bayesian networks

A BN was used to estimate the probability of a patient having PsA given their clinical factors measured prior to diagnosis. A BN is a probabilistic graphical model that can be visualised using a directed acyclic graph. Variables are represented by nodes in the graph, and conditional dependencies between variables by edges. An edge pointing from one variable (the parent) to another (the child) means that the probability distribution of the child node depends on the value of the parent node [17].

The graph structure of the BN was learned from the data, allowing us to discover complex interactions between the variables. The graph can also be used to perform variable selection by identifying the nodes in the Markov blanket of PsA. The Markov blanket contains all parents, children, and parents of the children of the PsA node. When performing inference, only the variables in the Markov blanket are necessary [17].

To learn the structure of a BN, we used the score-based structure learning algorithm Tabu search to maximise the Bayesian information criterion score. We compared the graph learned solely using Tabu search to an approach based on the Markov blanket Bayesian classifier (MBBC) [18]. The MBBC splits the structure learning process into 2 stages, where we begin by only learning parents and children of PsA, before fixing these edges and learning the rest of the graph as usual. This 2-stage strategy prioritises the fit of the graph around the PsA node, which may improve classification performance.

To ensure our confidence in the learned edges, we also performed model averaging. We repeated structure learning on 200 bootstrapped samples of the data and selected the edges that appeared in at least 85% of the resulting graphs.

As our variables are discrete, each node has an associated conditional probability table dependent on the values of that node's parents in the graph. We fit these parameters to the data using the Bayesian posterior estimator with a uniform prior.

Predictions using the BN were calculated using likelihood weighting, a Monte Carlo posterior inference procedure [19].

Comparator models

We fit 2 other models from the family of BNs: a naïve Bayes (NB) and a tree augmented network (TAN). A NB is a simple network where the PsA node is the parent of all other nodes, and the explanatory nodes are assumed to be independent. A TAN slightly relaxes this restriction to allow each explanatory node to depend on one other in a tree structure. The NB and TAN are commonly used for classification tasks as by design all nodes are used for prediction. We also fit a RF, a commonly used and straightforward to implement machine learning model to use as a performance benchmark.

Model evaluation

We evaluated the predictive probabilities of the models in terms of discrimination and calibration. We performed 5-fold cross validation, stratifying the data to ensure that each fold contained an equal prevalence of PsA patients.

Discrimination measures how well the model can differentiate between patients with and without PsA. The overall discrimination of the models was measured using the area under the receiver operating characteristic (ROC) curve (AUC), and the area under the precision-recall curve (PRAUC). The PRAUC has been suggested to be more appropriate than AUC in the context of imbalanced data as it focuses on positive class predictions [20]. A noninformative model would produce an AUC of 0.5, and a PRAUC equal to the prevalence of the positive class [21], which in our data is around 0.02. In addition, 95% CIs were estimated using bootstrapping.

Calibration assesses how well the predicted probabilities reflect the true probability of developing PsA. Calibration was measured using calibrations plots, created by binning the predicted probabilities into 10 bins and plotting the observed probabilities against the mean predicted probabilities. Where poor calibration was observed, beta calibration [22] was included as a postprocessing step. A beta model is fit to map the predicted probabilities to the observed probabilities on the training data, which is then used to calibrate the predictions on the test data.

Model interpretation

Permutation variable importance was used to identify which variables were the most influential for model predictions. A variable is randomly permutated, and the effect this has on the model performance is calculated [23]. The more important the variable is, the larger the decrease in performance caused by the permutation.

Relative risks (RRs) for each variable were calculated for the BN. To calculate the RR for a BN, we used conditional probability queries to calculate the absolute risk of PsA, conditional only

on each level of the variable of interest. One level of the variable was chosen as the reference, and the relative risk with respect to this level was calculated as the ratio of the absolute risks.

Software

Analysis was performed using RStudio [24] with R version 4.2.0 [25]. BNs were developed using the *bnlearn* package, version 5.0 [19]. RFs were fitted using Python version 3.11.5 and the scikit-learn package [26] version 1.3.0. AUC and PRAUC were calculated using the *MLMetrics* package [27] version 1.1.3.

RESULTS

Study population

We identified 122,330 patients in the CPRD with incident psoriasis who fulfilled the study inclusion and exclusion criteria. The overall process of identifying the study population is summarised in Figure 1. Of these patients, 14,340 patients had their psoriasis diagnosis backdated, with a median (IQR) backdating duration of 310 (76, 997) days. The median (IQR) follow-up duration of psoriasis patients was 5.81 (2.64, 10.1) years. We identified 2460 patients who went on to develop PsA, with 388 (13.7%) having their PsA diagnosis backdated on the basis of a csDMARD prescription, with a median (IQR) backdating of 206 (61, 730) days. We excluded 55 PsA patients from backdating via methotrexate because of evidence that their prescription may have been associated with a dermatology referral. The median (IQR) time between psoriasis and PsA diagnosis was 2.87 (0.912, 6.44) years. Study population demographics are summarised in the Supplementary Material (Supplementary Table S3).

Bayesian networks

The Markov blanket of the PsA node in the BNs learned using Tabu search (Tabu) and the MBBC procedure (MBBC) are shown in Figure 2. We found that the network learned by MBBC included an extended subset of variables compared to the network learned by Tabu. Tabu did not contain any demographic or blood test variables, whereas MBBC included age, BMI, and alcohol status, as well as CRP and PV test results. In addition, MBBC included psoriasis treatment, and a much larger collection of musculoskeletal symptoms.

Predictive performance

The ROC and precision-recall curves averaged over the 5 cross validation sets are shown in Figure 3. The AUC and PRAUC, with 95% CIs, are shown in Table. Comparing PRAUC, the RF had the best performance, closely followed by the MBBC BN and TAN. MBBC outperforms the Tabu BN, showing that the larger subset of variables identified did result in a significant improvement in model performance. The worst PRAUC was found by the NB.

The calibration of the models was assessed visually before and after the application of beta calibration (Supplementary Fig). The NB, TAN and MBBC all benefitted from beta calibration, as they tended to overestimate the probability of developing PsA. After scaling, the models all show reasonable calibration. Tabu is largely unchanged, showing good calibration before and after scaling. The RF, however, did not benefit from beta calibration, with calibration worsened by the procedure.

For the remainder of this section, we prioritise the results from the RF and MBBC BN, which were the 2 best performing models in terms of PRAUC.

Model interpretation

The most important variables were identified using cross-validated permutation variable importance, measured by the reduction in PRAUC. The combined top 10 variables for the RF and MBBC BN are shown in Figure 4. The RRs for the MBBC BN are shown in Figure 5.

We saw much agreement between which variables are considered important by the 2 models, with only nonspecific swelling not making the top 10 of the RF, and smoking status not included in the MBBC BN. The most important variable for both models was psoriasis duration, and the RRs reveal the highest probability associated within the first year from psoriasis diagnosis. We found an increased risk of PsA in patients with a prescription for an NSAID, as well as those with the musculoskeletal symptoms nonspecific arthritis, nonspecific arthralgia, and nonspecific swelling. A CRP blood test was an

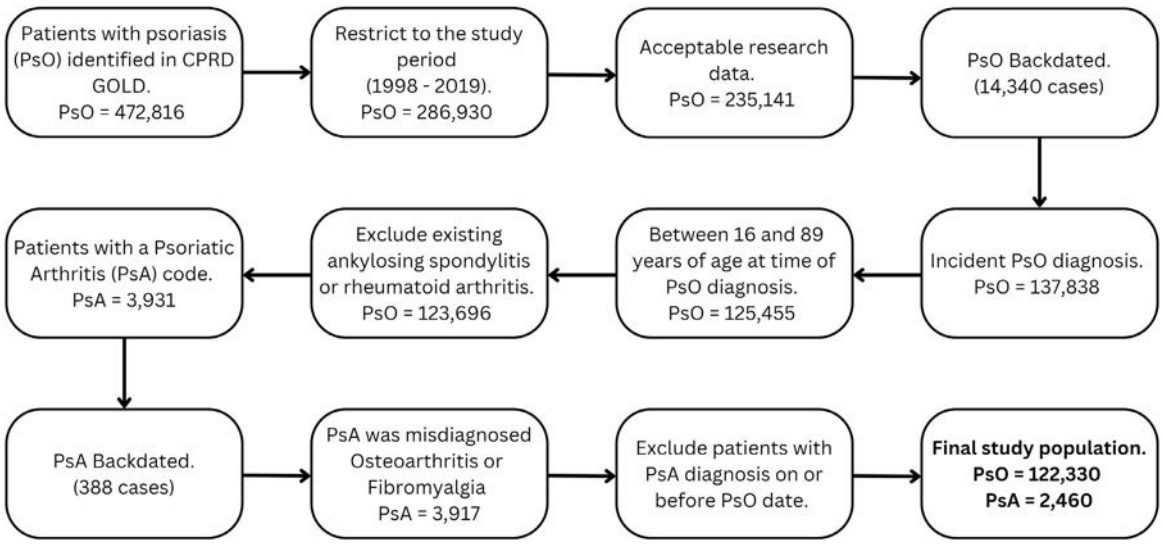


Figure 1. Flow diagram showing how the study population was derived. Boxes show inclusion or exclusion criteria and the number of patients that fulfil the criteria. Arrows between boxes point in the direction of the sequential ordering of the procedure. CPRD, Clinical Practice Research Datalink.

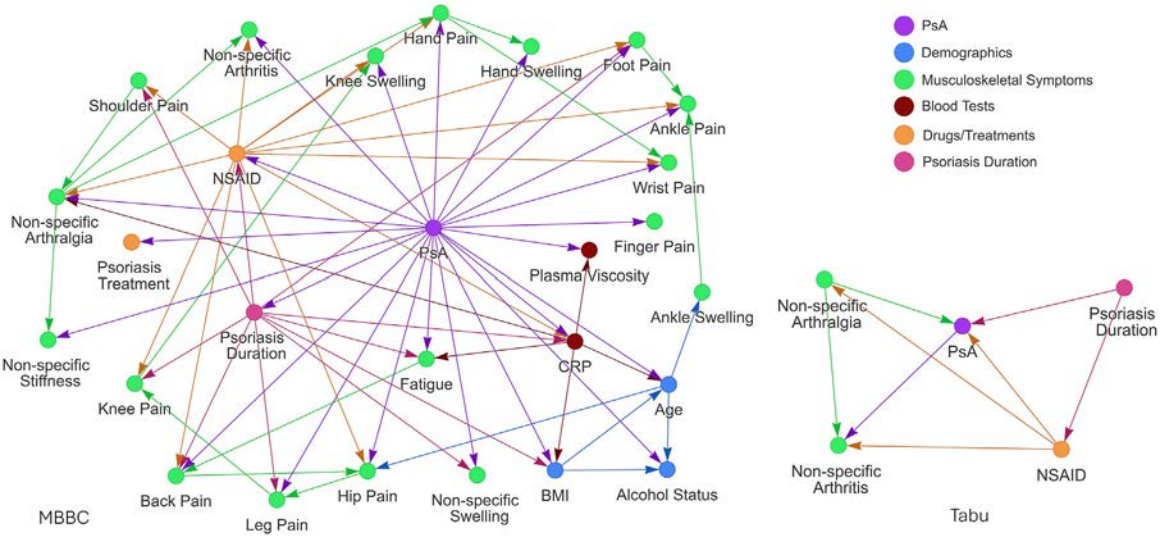


Figure 2. The Markov blankets of the PsA node in the Bayesian networks learned using MBBC (left) and Tabu (right) search. Arrows represent conditional dependencies between variables in the network. BMI, body mass index; CRP, C-reactive protein; MBBC, Markov blanket Bayesian classifier; NSAID, nonsteroidal anti-inflammatory drugs; PsA, psoriatic arthritis.

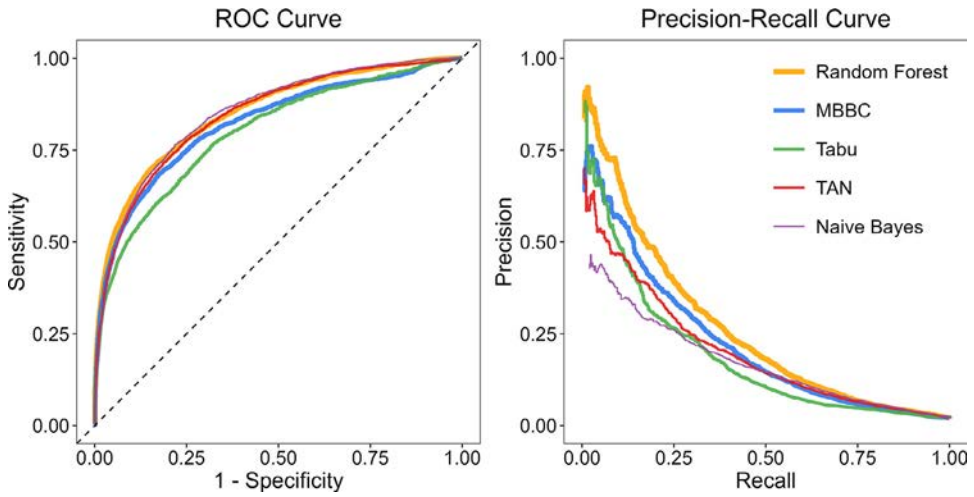


Figure 3. Five-fold cross-validated ROC (left) and precision-recall curves (right). MBBC, Markov blanket Bayesian classifier; TAN, tree augmented network; ROC, receiver operating characteristic.

important indicator of PsA in the models. Compared to patients with no CRP test, we found increased probabilities in patients who did have a CRP blood test, even if the result was normal, with the highest probability associated with a high CRP. For demographic variables, age, BMI, and alcohol status were all important. Risk was elevated in patients in age groups between 30 and 60 years as well as those with a BMI above the normal range.

Table

Five-fold cross-validated AUC and PRAUC estimates and lower and upper bounds of bootstrapped 95% CIs

	AUC			PRAUC		
Model	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
Random forest	0.851	0.842	0.859	0.261	0.242	0.280
Bayesian network (MBBC)	0.823	0.813	0.833	0.221	0.204	0.240
Tree augmented network	0.845	0.836	0.853	0.198	0.182	0.215
Bayesian network (Tabu)	0.799	0.789	0.809	0.185	0.168	0.201
Naïve Bayes	0.851	0.843	0.859	0.172	0.159	0.186

AUC, area under the receiver operating characteristic curve; MBBC, Markov blanket Bayesian classifier; PRAUC, area under the precision-recall curve.

DISCUSSION

In this study, we developed a BN to identify psoriasis patients at higher risk of PsA in a large UK primary care cohort from the CPRD. Using clinical variables measured prior to diagnosis, we identified important indicators of PsA in primary care.

We built upon analysis reported in Green et al [13] by taking advantage of a longer period of data collection within the CPRD and using a wider variety of BNs. We achieved a better predictive performance, improving the AUC from 0.73 to over 0.80. We included more clinical predictors, such as NSAID prescriptions and CRP blood tests, which Green et al [13] suggested were potential avenues of further study.

We found that an NSAID prescription and having had a CRP blood test were both important variables in our models, corroborating findings in a recent model by Shapiro et al [12]. Prescriptions and blood tests are variables determined by a patient's interactions with primary care, as GPs order them as a response

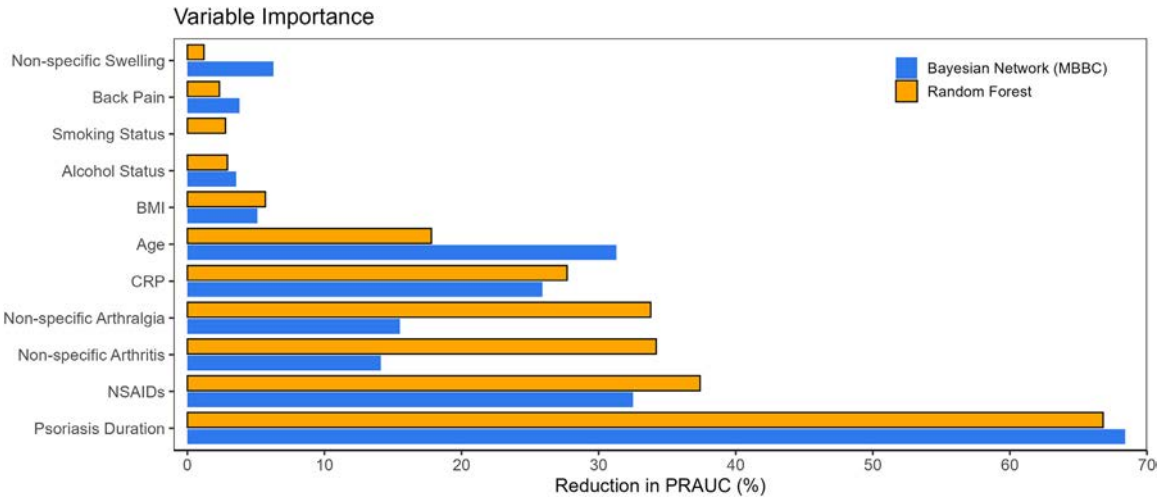


Figure 4. Permutation variable importance for the Bayesian network and random forest models, measured as the reduction in PRAUC. More important variables result in a greater reduction in PRAUC after permutation. BMI, body mass index; CRP, C-reactive protein; MBBC, Markov blanket Bayesian classifier; NSAID, nonsteroidal anti-inflammatory drugs; PRAUC, area under the precision-recall curve.

to symptoms or signs, which could be a result of undiagnosed PsA. Patients with an NSAID prescription had an elevated risk of being diagnosed with PsA in our model, with NSAIDs being a common medication for inflammatory arthritis symptoms. However, in clinical practice, where inflammatory musculoskeletal symptoms can be verified directly with the patient, the presence of NSAIDs may be less informative. An elevated CRP level was also associated with PsA; however, we also found that a normal CRP blood test result was associated with an increased risk compared to no blood test at all. A potential explanation for the association of normal CRP levels with PsA is that a GP may request a CRP test in response to inflammatory arthritis that is yet to be diagnosed as PsA. However, PsA does not always result in an elevated CRP level [28].

We identified an unexpected increase in PsA probability within the first year from psoriasis diagnosis. Tillett et al [29] found a peak in synchronous onset of PsA and psoriasis (PsO) in both the CPRD and a secondary care cohort, with PsO being detected and recorded in parallel with a diagnosis of PsA being made. If this is the case in our data, the inclusion criteria may be being violated as we require incident cases of psoriasis with no pre-existing PsA. Therefore, we suggest that the association with more recent onset of psoriasis is an artefact of the data, rather than a true increase in risk.

A strength of this study is that we have used an interpretable BN. When compared to a RF, the BN offered a viable alternative in terms of performance. We argue that the small reduction in performance shown by the BN is offset by its superior interpretability, making it more suitable for clinical reasoning [30]. The graphical structure shows how the variables in the model interact with one another, allowing clinicians a better understanding on how the model is operating. If the model and its structure are consistent with their own knowledge of the disease, then the model is more likely to be trusted by the clinician. Explainable predictions are also easier to combine with a clinician's own ideas, allowing them to supplement the decision-making process. Additionally, as BNs allow inference using any subset of the variables in the model, they can be used even when information on some of the variables are missing.

Another strength of our study is that EHRs provide a large sample size for fitting complex models. Eder et al [10] recognised that the PRESTO model was limited by the small sample size, which may have limited model complexity and the inclusion of some risk factors. However, the use of EHRs does have limitations. As PsA often goes undiagnosed [31,32], there is some uncertainty in our data regarding whether patients actually are PsA free. This is a limitation when using primary care data, whereas in secondary care cohorts the presence and absence of PsA can be verified. Furthermore, as PsA is primarily diagnosed in secondary care, there may be patients diagnosed with PsA whose primary care record has not been updated to reflect this. We accounted for lag between secondary and primary care diagnosis by backdating a patients diagnosis date using data on prescriptions and referrals, but lack of this information was common. The use of primary care data has other limitations. As data are not collected with research in mind, clinical observations investigated may go unrecorded by GPs if less relevant to the consultation at the time, and data are only recorded when a patient makes a visit to primary care. Therefore, patients may have symptoms that are not reflected in the data.

Because of these factors, our external validity outside of primary care is likely to be limited, and whether these models can be of use as a screening tool requires further study. From the precision-recall curve (Fig 3), we can also see that the classification performance of all models investigated exhibited poor precision, particularly at higher levels of recall. If applied to real life health care scenarios, we are likely limited by an acceptable level of precision, as there are often limited resources for screening and diagnosis. Existing methods such as the Psoriasis Epidemiology Screening Tool (PEST) have shown similar limitations. In a screening of a primary care psoriasis population, PEST achieved a precision of 0.146 when using the usual positive cut-off score of 3 or more [33]. For clinical prediction tools, calibration is also important, and we made use of beta calibration to help improve calibration of the models which were not naturally well calibrated.

In future work, we aim to develop a BN to produce dynamic future risk estimates, allowing us to predict the risk of

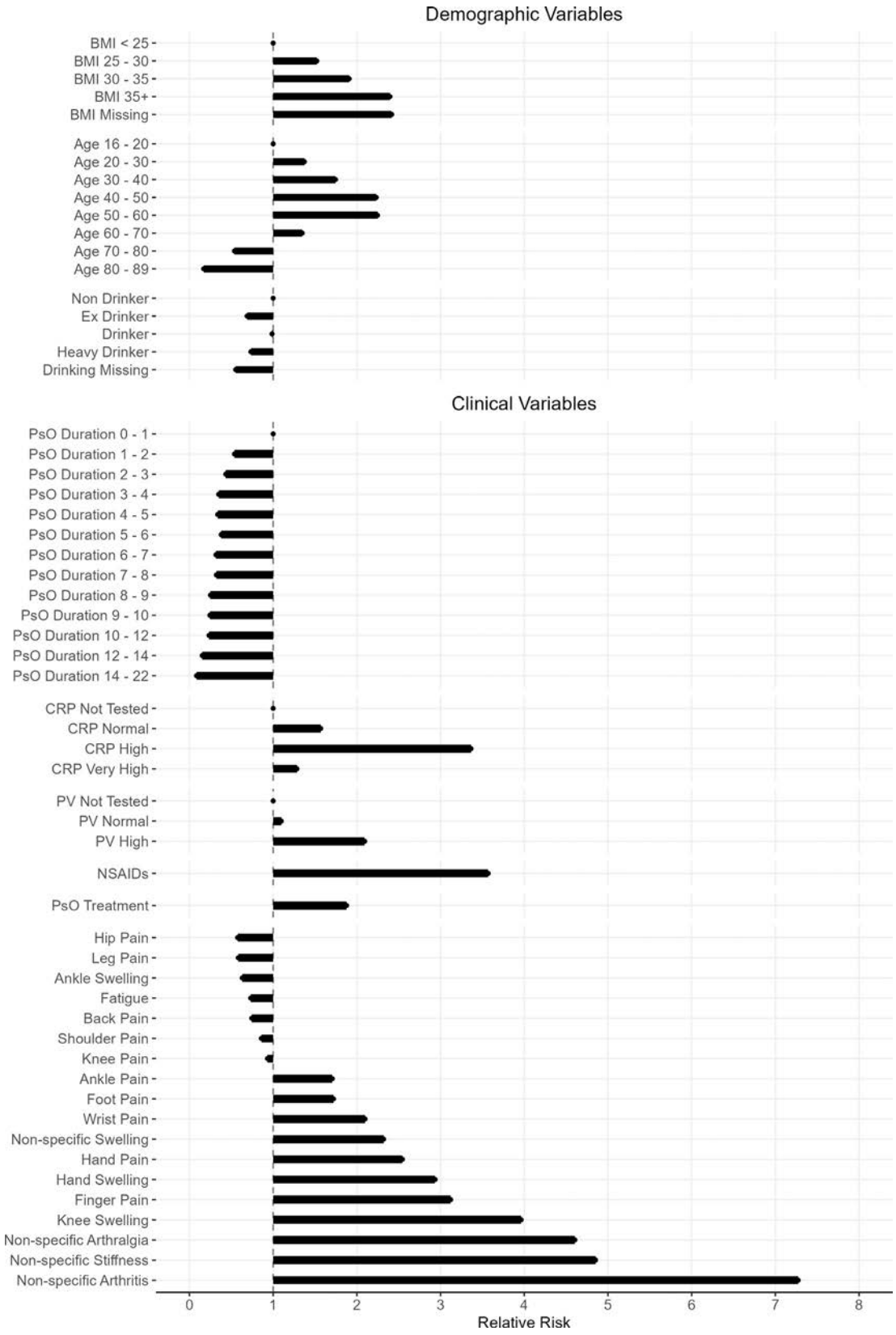


Figure 5. Relative risks calculated using the MBBC Bayesian network model. BMI, body mass index; CRP, C-reactive protein; MBBC, Markov blanket Bayesian classifier; NSAIDs, nonsteroidal anti-inflammatory drugs; PsO, psoriasis; PV, plasma viscosity.

developing PsA as symptoms evolve over time and evaluate its ability to facilitate the earlier identification of PsA.

Competing interests

AR reports financial support was provided by UCB Biopharma SRL. NM reports grant funding from UCB Biopharma SRL and speaking and lecture fees from Janssen Pharmaceuticals Inc. WT reports consulting or advisory and speaking and lecture fees from AbbVie Inc and Amgen Inc; consulting or advisory to Bristol Myers Squibb, GSK, and Ono Pharma UK Ltd; funding grants and speaking and lecture fees from Celgene; consulting or advisory, funding grants, speaking and lecture fees, and travel reimbursement from Eli Lilly and Company, Janssen Pharmaceuticals Inc, Novartis, Pfizer, and UCB Biopharma SRL; and speaking and lecture fees from Merk Sharp & Dohme. The other author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the University of Bath's Research Computing Group (doi.org/10.15125/b6cd-s854) for their support in this work.

Funding

This investigator-initiated study was financially supported by UCB Biopharma SRL.

Patient consent for publication

Not applicable.

Ethics approval

The use of Clinical Practice Research Datalink data for this study was approved by the Research Data Governance Process from the Medicines and Healthcare products Regulatory Agency (Protocol ID 21_000578).

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

Data used for this study is not publicly available. Data from the Clinical Practice Research Datalink is under licence from the UK Medicines and Healthcare Products Regulatory Agency, which restricts the sharing of data to persons outside the study protocol.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.051.

Orcid

Alexander Rudge: http://orcid.org/0009-0001-5419-6286
Theresa Smith: http://orcid.org/0000-0002-7085-3864

REFERENCES

[1] Alinaghi F, Calov M, Kristensen LE, Gladman DD, Coates LC, Jullien D, et al. Prevalence of psoriatic arthritis in patients with psoriasis: a systematic review and meta-analysis of observational and clinical studies. J Am Acad Dermatol 2019;80(1) 251-65.e19. doi: 10.1016/j.jaad.2018.06.027.
[2] Coates LC, Helliwell PS. Psoriatic arthritis: state of the art review. Clin Med (Lond) 2017;17(1):65–70. doi: 10.7861/clinmedicine.17-1-65.
[3] Eder L, Polachek A, Rosen CF, Chandran V, Cook R, Gladman DD. The development of psoriatic arthritis in patients with psoriasis is preceded by a period of nonspecific musculoskeletal symptoms: a prospective cohort study. Arthritis Rheumatol 2017;69(3):622–9. doi: 10.1002/art.39973.
[4] Scher JU, Ogdie A, Merola JF, Ritchlin C. Preventing psoriatic arthritis: focusing on patients with psoriasis at increased risk of transition. Nat Rev Rheumatol 2019;15(3):153–66. doi: 10.1038/s41584-019-0175-0.
[5] Zabotti A, Fagni F, Gossec L, Giovannini I, Sticherling M, Tullio A, et al. Risk of developing psoriatic arthritis in psoriasis cohorts with arthralgia: exploring the subclinical psoriatic arthritis stage. RMD Open 2024;10(2):e004314. doi: 10.1136/rmdopen-2024-004314.
[6] Ogdie A, Weiss P. The epidemiology of psoriatic arthritis. Rheum Dis Clin North Am 2015;41(4):545–68. doi: 10.1016/j.rdc.2015.07.001.
[7] Ogdie A, Rozycki M, Arndt T, Shi C, Kim N, Hur P. Longitudinal analysis of the patient pathways to diagnosis of psoriatic arthritis. Arthritis Res Ther 2021;23(1):252. doi: 10.1186/s13075-021-02628-2.
[8] Tillett W, Jadon D, Shaddick G, Cavill C, Korendowych E, de Vries CS, et al. Smoking and delay to diagnosis are associated with poorer functional outcome in psoriatic arthritis. Ann Rheum Dis 2013;72(8):1358–61. doi: 10.1136/annrheumdis-2012-202608.
[9] Haroon M, Gallagher P, FitzGerald O. Diagnostic delay of more than 6 months contributes to poor radiographic and functional outcome in psoriatic arthritis. Ann Rheum Dis 2015;74(6):1045–50. doi: 10.1136/annrheumdis-2013-204858.
[10] Eder L, Lee KA, Chandran V, Widdifield J, Drucker AM, Ritchlin C, et al. Derivation of a multivariable psoriatic arthritis risk estimation tool (PRESTO): a step towards prevention. Arthritis Rheumatol 2023. doi: 10.1002/art.42661.
[11] Ogdie A, Harrison RW, McLean RR, Lin TC, Lebwohl M, Strober BE, et al. Prospective cohort study of psoriatic arthritis risk in patients with psoriasis in a real-world psoriasis registry. J Am Acad Dermatol 2022;87(6):1303–11. doi: 10.1016/j.jaad.2022.07.060.
[12] Shapiro J, Getz B, Cohen SB, Jenudi Y, Underberger D, Dreyfuss M, et al. Evaluation of a machine learning tool for the early identification of patients with undiagnosed psoriatic arthritis - A retrospective population-based study. J Transl Autoimmun 2023;7:100207. doi: 10.1016/j.jtauto.2023.100207.
[13] Green A, Tillett W, McHugh N, Smith T, PROMPT Study Group. Using Bayesian networks to identify musculoskeletal symptoms influencing the risk of developing psoriatic arthritis in people with psoriasis. Rheumatology (Oxford) 2022;61(2):581–90. doi: 10.1093/rheumatology/keab310.
[14] Polotskaya K, Muñoz-Valencia CS, Rabasa A, Quesada-Rico JA, Orozco-Beltrán D, Barber X. Bayesian networks for the diagnosis and prognosis of diseases: a scoping review. Mach Learn Knowl Extract 2024;6(2):1243–62. doi: 10.3390/make6020058.
[15] Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol 2015;44(3):827–36. doi: 10.1093/ije/dyv098.
[16] Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. Pharmacoepidemiol Drug Saf 2005;14(7):443–51. doi: 10.1002/pds.1115.
[17] Scutari M, Denis J-B. Bayesian Networks: With Examples in R. 2nd ed. Chapman and Hall/CRC; 2021. doi: 10.1201/9780429347436.
[18] Madden MG. Evaluation of the performance of the Markov blanket Bayesian classifier algorithm. arXiv: cs/0211003 [Preprint]. 2002 [cited 6th September 2024]. Available from: https://doi.org/10.48550/arXiv.cs/0211003
[19] Scutari M. Learning Bayesian networks with the bnlearn R Package. Journal of Statistical Software 2010;35(3). doi: 10.18637/jss.v035.i03.
[20] Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10(3):e0118432. doi: 10.1371/journal.pone.0118432.
[21] Fernández A, García S, Galar M, et al. Learning from Imbalanced Data Sets. Springer Cham; 2018. doi: 10.1007/978-3-319-98074-4.
[22] Kull M, Filho TS, Flach P. Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers. In: Aarti S,

Jerry Z, editors. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research; 2017. p. 623–31.
[23] Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd ed.; 2022. [cited 6th September 2024]. Available from https://christophm.github.io/interpretable-ml-book/.
[24] RStudio Team. RStudio: Integrated Development for R; 2020. Available from http://www.posit.co
[25] R Core Team. R. a language and environment for statistical computing. R Foundation for Statistical Computing; 2024. Available from https://www.R-project.org/.
[26] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[27] Yan Y. MLmetrics: Machine Learning Evaluation Metrics. R Package 1.1.3 version; 2025 https://github.com/yanyachen/mlmetrics.
[28] Gialouri CG, Evangelatos G, Pappa M, Karamanakos A, Iliopoulos A, Tektonidou MG, et al. Normal C-reactive protein in active psoriatic arthritis: results from real-world clinical practice. Ther Adv Musculoskelet Dis 2022;14. doi: 10.1177/1759720X221122417.
[29] Tillett W, Charlton R, Nightingale A, Snowball J, Green A, Smith C, et al. Interval between onset of psoriasis and psoriatic arthritis comparing the UK Clinical Practice Research Datalink with a hospital-based cohort. Rheumatology (Oxford) 2017;56(12):2109–13. doi: 10.1093/rheumatology/kex323.
[30] Lu SC, Swisher CL, Chung C, Jaffray D, Sidey-Gibbons C. On the importance of interpretable machine learning predictions to inform clinical decision making in oncology. Front Oncol 2023;13:1129380. doi: 10.3389/fonc.2023.1129380.
[31] Mease PJ, Gladman DD, Papp KA, Khraishi MM, Thaçi D, Behrens F, et al. Prevalence of rheumatologist-diagnosed psoriatic arthritis in patients with psoriasis in European/North American dermatology clinics. J Am Acad Dermatol 2013;69(5):729–35. doi: 10.1016/j.jaad.2013.07.023.
[32] Ibrahim G, Waxman R, Helliwell PS. The prevalence of psoriatic arthritis in people with psoriasis. Arthritis Rheum 2009;61(10):1373–8. doi: 10.1002/art.24608.
[33] Helliwell PS, Coates LC, Ransom M, Brown ST, Packham J, Weddell J, et al. The comparative performance of three screening questionnaires for psoriatic arthritis in a primary care surveillance study. Rheumatology (Oxford) 2024;63(4):991–8. doi: 10.1093/rheumatology/kead310.

Ann Rheum Dis 84 (2025) 584–593

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Axial spondyloarthritis

Impact of clinical subtype and sex on first-line biologic therapy discontinuation in axial spondyloarthritis

Patricia Remalante-Rayco[1,2], Emmanuel S Baja[2,3], Zeynep Baskurt[4], Tina Chim[5], Carlo Irwin A Panelo[2], Evelyn Osio-Salido[6], Robert D Inman[1,7], Leonila F Dans[2,8], Nigil Haroon[1,7,*]

[1] *Schroeder Arthritis Institute, Toronto Western Hospital, Spondylitis Program, University Health Network, Toronto, ON, Canada*
[2] *Department of Clinical Epidemiology, College of Medicine, University of the Philippines Manila, Manila, Philippines*
[3] *Institute of Clinical Epidemiology, National Institutes of Health, University of the Philippines Manila, Manila, Philippines*
[4] *Department of Biostatistics, University Health Network, Toronto, ON, Canada*
[5] *Schroeder Arthritis Institute, Toronto Western Hospital, University Health Network, Toronto, ON, Canada*
[6] *Division of Rheumatology, Department of Medicine, Philippine General Hospital, University of the Philippines Manila, Manila, Philippines*
[7] *Department of Medicine, Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada*
[8] *Department of Pediatrics, College of Medicine, University of the Philippines Manila, Manila, Philippines*

ARTICLE INFO

Article history:
Received 14 August 2024
Received in revised form 19 December 2024
Accepted 20 December 2024

ABSTRACT

Objectives: To estimate the main and interaction effects of axial spondyloarthritis (axSpA) subtype and sex on first biologic disease-modifying antirheumatic drug (bDMARD) discontinuation.
Methods: This retrospective cohort study included nonradiographic axSpA (nr-axSpA) and radiographic axSpA (r-axSpA) patients initiating tumour necrosis factor or interleukin-17 inhibitors. Modified Poisson regressions were used to estimate risk ratios (RRs) for the association of subtype and sex with discontinuation, adjusting for baseline covariates. Interaction was assessed using the relative excess risk due to interaction (RERI) and ratio of RRs. In addition, bDMARD survival rates were analysed using Kaplan–Meier curves.
Results: Among 469 patients, 64% discontinued their first bDMARD. Nr-axSpA (RR, 1.80; 95% CI, 1.26-2.59) and female sex (RR, 1.49; 95% CI, 1.081-2.045) were significantly associated with discontinuation. Positive interaction trends between subtype and sex were observed on additive (RERI 0.49, 95% CI, −0.78 to 1.75) and multiplicative (RR ratio, 1.05; 95% CI, 0.55-2.03) scales, though not statistically significant. Nr-axSpA females had twice the discontinuation risk of r-axSpA males (hazard ratio, 2.30; 95% CI, 1.68-3.15, $P < .001$). bDMARD survival over 20 years was significantly lower in nr-axSpA and female patients.
Conclusions: Nr-axSpA and female patients face a significantly higher risk of bDMARD discontinuation and shorter bDMARD survival. Although the combined effect of subtype and sex trended higher, it was not statistically significant. These findings underscore the need to address potential treatment challenges in female nr-axSpA patients.

*Correspondence to Dr Nigil Haroon, Schroeder Arthritis Institute, Toronto Western Hospital, Spondylitis Program, University Health Network, Toronto, ON, Canada.
E-mail address: Nigil.Haroon@uhn.ca (N. Haroon).
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.01.007

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Lower response rates to biologic disease-modifying antirheumatic drugs (bDMARDs) among female patients has been reported. However, there is a paucity of studies exploring the interaction between axial spondyloarthritis (axSpA) subtype (nonradiographic axSpA [nr-axSpA] or radiographic axSpA) and patient sex on first-line bDMARD discontinuation.

WHAT THIS STUDY ADDS

- The nr-axSpA subtype and female sex confer a significantly higher risk of first-bDMARD discontinuation and lower bDMARD survival rates. These observed effects were robust to unmeasured confounding.
- Although not achieving statistical significance, the risk of bDMARD discontinuation trended higher when both the nr-axSpA subtype and female sex were present.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- We highlight the poorer treatment outcomes and unmet needs of female nr-axSpA patients. This observation may guide research exploring underlying biological mechanisms and inform the development of tools to accurately measure disease burden and outcomes in this subgroup.
- Larger studies are needed to validate our finding of a potential interaction effect between subtype and sex on bDMARD discontinuance.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a chronic inflammatory rheumatic disease affecting the sacroiliac joints (SIJ) and spine that can lead to impaired function and reduced quality of life. Over the past 15 years, axSpA has been viewed as a disease spectrum encompassing 2 subtypes: radiographic axSpA (r-axSpA) and nonradiographic axSpA (nr-axSpA). Nr-axSpA is often considered an early form of axSpA because bone marrow oedema indicative of active inflammation on magnetic resonance imaging (MRI) has been shown to precede structural damage in the SIJs and spine [1,2]. The idea of a nr-axSpA subtype has allowed for earlier disease recognition and intervention. However, its being an early form of r-axSpA remains contentious as only a subset develops r-axSpA [3]. Furthermore, SIJ bone marrow oedema can be nonspecific and may be seen in mechanical back pain patients, runners, postpartum women, and even healthy individuals [4]. Bone marrow oedema may likewise be missed in up to 30% of patients with spondyloarthritis (SpA) features and a positive human leukocyte antigen B27 (HLA-B27) test [5].

Some variability has been found in the disease presentation of r-axSpA and nr-axSpA. R-axSpA typically has an earlier symptom onset and longer symptom duration, worse spinal mobility and function, higher levels of the inflammatory markers C-reactive protein (CRP) and erythrocyte sedimentation rate, and a higher prevalence of HLA-B27 in most studies [6–10]. However, disease activity, physical and mental health, and quality of life are comparable for the 2 subtypes [10–13].

Notably, recent evidence indicated poorer outcomes and lower response rates to tumour necrosis factor inhibitors (TNFis) in female nr-axSpA patients. Females with nr-axSpA less often achieved low Ankylosing Spondylitis Disease Activity Scores (ASDASs) and inactive disease at 1 year from diagnosis than males, a difference not observed in r-axSpA [14]. A Swiss cohort found that fewer females with nr-axSpA achieved significant Assessment in SpondyloArthritis international Society 40 (ASAS40) treatment response and ASDAS clinically important improvement after 1 year of TNFi therapy [6]. Pooled data from at least 10 randomised controlled trials (RCTs) revealed that male axSpA patients had higher ASAS40 responses to biologic disease-modifying antirheumatic drugs (bDMARD) compared with females, with a more pronounced sex disparity favouring males in nr-axSpA [15]. However, the impact of these findings has not been investigated from the perspective of bDMARD discontinuation or survival, which is a proxy composite measure reflecting bDMARD effectiveness, safety, and tolerability in the real world.

Although women have been consistently shown to have increased bDMARD discontinuance, the difference conferred by subtype has been conflicting. Given the higher prevalence of nr-axSpA and poorer adherence or treatment outcomes among females, understanding the interaction between subtype and sex on bDMARD discontinuation is crucial. We aimed to estimate the direct effects of axSpA subtype and sex and their interaction on all-cause discontinuation of the first-line TNFi or interleukin-17 inhibitor (IL-17i). As secondary objectives, we compared bDMARD survival rates and the hazard of first bDMARD discontinuation by subtype, sex, and the combination of both.

METHODS

Study design and population

This was a retrospective cohort study of prospectively collected data from a cohort of axSpA patients followed at the Spondylitis Program of the Schroeder Arthritis Institute. Patients were followed every 6 to 12 months as per protocol. We included patients who first received a diagnosis of axSpA by a rheumatologist and were thereafter classified according to the 2009 ASAS classification criteria with (1) a baseline visit between June 1, 2003, and June 30, 2023; (2) defined start dates for their first TNFi or IL-17i; and (3) at least 1 follow-up visit. The exclusion criteria were (1) commencement of the first bDMARD prior to the baseline visit, (2) receipt of a non-TNFi or non-IL-17i first-line bDMARD, (3) missing first-bDMARD start dates, or (4) coexisting non-SpA inflammatory arthritis or systemic autoimmune rheumatic disease. Subtype classification was facilitated using pelvic radiograph readings by 2 experienced local readers and 1 adjudicating reader in the case of discrepancy.

Outcomes, exposures, and covariates

The predetermined primary outcome of interest was the risk of all-cause first-line bDMARD discontinuation, measured as the risk ratio (RR), with subtype and sex as the key exposures under investigation. The RR was used as the primary effect measure because it directly quantifies the cumulative probability of an event over time in exposed versus unexposed groups. Compared with the odds ratio, the RR offers a more accurate estimate of the effect size when the outcomes are not rare [16–19]. Additionally, it reflects actual changes in risk rather than potentially inflated estimates and is thus more relevant in clinical decision making, risk communication, and evaluation of health interventions.

The secondary outcomes were (1) first-line bDMARD survival rate, which refers to the proportion of patients discontinuing the first bDMARD at 1, 5, 10, and 15 years; and (2) the hazard ratio (HR) for bDMARD discontinuance, which represents the risk of discontinuance at any given point in time during the study

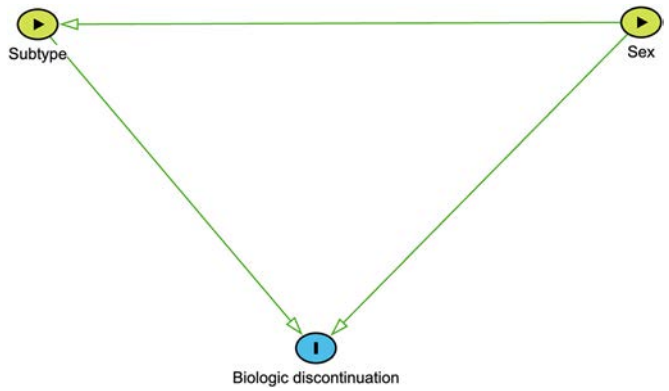


Figure 1. Simplified graph of the relationships among subtype, sex, and biologic discontinuation.

period. These were calculated using the start and stop dates documented in the protocol. Observations were censored for loss to follow-up, death, or the end of the observation period (June 30, 2023), whichever came first. Temporary treatment interruptions were allowed according to the treatment cycle of the first bDMARD: <1 month for adalimumab, certolizumab pegol, and etanercept; and <2 months for golimumab, infliximab, secukinumab, and ixekizumab.

AxSpA subtype (nr-axSpA versus r-axSpA) and sex (female versus male) were the exposures of interest. Nr-axSpA and female sex served as the test variables, whereas r-axSpA and male sex were treated as control variables. We selected the patient's subtype at bDMARD initiation given that baseline factors associated with bDMARD discontinuance were of interest. To identify our covariates for the estimation of the direct effect of subtype and sex on bDMARD discontinuation, we constructed a directed acyclic graph (DAG) (Fig 1, Supplementary Fig 1). Understanding the direct effect of these exposures allows us to determine whether observed differences in the outcome arise from mechanisms related to intrinsic differences between subtypes and sexes. The DAG was created with the DAGitty 3.0 programme [20] using covariates determined *a priori* from literature review and knowledge of the subject matter (Supplementary Fig 1). The minimal sufficient adjustment set included 15 covariates: age, symptom duration, Bath Ankylosing Spondylitis Disease Activity Index, Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Metrology Index, CRP levels, and comorbidity counts as continuous variables as well as HLA-B27, ever smoking, ever fibromyalgia, ever use of a conventional synthetic disease-modifying antirheumatic drug (csDMARD), and ever presence of peripheral arthritis, uveitis, inflammatory bowel disease (IBD), and psoriasis as dichotomous variables. Fibromyalgia was treated as a comorbidity of special interest and included as a separate covariate. Within this framework, symptom duration, HLA-B27, age, and smoking were identified as confounders, whereas the remaining covariates functioned as mediators. Adjusting for both confounders and mediators enabled us to isolate and estimate the direct effects of subtype and sex. All variables were determined at baseline, defined as 3 months before or after starting the first bDMARD.

Statistical analyses

Main and interaction effects of subtype and sex on bDMARD discontinuation

We used modified Poisson (Poisson regression with robust error variance) to obtain RRs and 95% CIs [19]. The duration of bDMARD use was included as an offset term in the Poisson regression models to account for varying lengths of observation periods across patients. We performed modified Poisson regression first with subtype or sex alone as the independent variable, and then with adjustment for the 15 covariates in our minimal sufficient adjustment set, using bivariable and multivariable models, respectively. Collinearity was evaluated using the variance inflation factor (VIF), with a threshold of 10 for acceptable predictor correlation.

The relative excess risk due to interaction (RERI) and its 95% CI were estimated to assess additive interaction between subtype and sex. A RERI > 0 suggests a stronger association between the 2 among those who discontinued bDMARDs, highlighting patient groups that may benefit from intervention [21]. Multiplicative interaction was measured using the ratio of RRs, with a ratio > 1 indicating a potential synergistic mechanism between subtype and sex. This information may be useful for predicting the risk of the outcome in patients with one or both exposures [22]. This analysis was performed using a modified version of the interactionR package version 0.1.7 [23].

To assess the effect of unmeasured confounding, expected values (E-values) were obtained. The E-value is the minimum strength of association on the RR scale that an unmeasured confounder must have with both the exposure and outcome variables to explain their association, conditional on measured covariates [24]. Interpretation of the E-value relies on context: E-values that are greater than the RR of known risk factors suggest a stronger association between the exposures of interest and the outcome and therefore robustness to unmeasured confounding [25]. E-values can be calculated for both the RR of the observed outcome and its 95% CI, specifically the CI limit closest to the null [25].

Survival analysis

Survival of the first bDMARD was compared between nr-axSpA and r-axSpA, males and females, and 4 patient subgroups (female nr-axSpA, male nr-axSpA, female r-axSpA, and male r-axSpA) using Kaplan–Meier plots and the log-rank test. HRs for all-cause bDMARD discontinuation were similarly estimated and compared to determine the rate of bDMARD discontinuance. To address the cohort bias related to the introduction of the nr-axSpA concept in 2009 and the historical bias from the arrival of newer bDMARDs, we conducted a predetermined subgroup analysis for 3 calendar periods based on the approval year of the bDMARD: (1) April 6, 2009, and earlier for etanercept, infliximab, and adalimumab; (2) April 7, 2009, to April 26, 2016, for golimumab and certolizumab pegol; and (3) April 27, 2016, to 2023 for secukinumab and ixekizumab.

Multiple imputation of missing values

Analysis of missing data patterns and clustering indicated that the data were not missing completely at random (MCAR). Therefore, multiple imputation by chained equations was employed. As the percentage of patients with any missing data approached 50%, we performed 50 imputations (each with 40 iterations to improve convergence). We imputed continuous variables using predictive mean matching and used logistic regression for binary variables. We assessed convergence through visual inspection of the mean and variance changes by iteration and dataset [26,27]. Model estimates were pooled across datasets according to Rubin's rules. The mice package version 3.16.0 was used for imputation [28]. Multiple imputation was applied to the regression analyses but not the survival analyses, as no variables in the latter had missing data.

All tests were 2-sided with a significance level of 0.05, applying corrections for multiple comparisons using the Hochberg method to balance Type I error control with statistical power. This method was chosen to account for our pre-specified and interdependent hypotheses. *P* value adjustment with the Hochberg method was performed using the p.adjust function in the stats R package. Data management and statistical analysis were done using the R Statistical Software (v4.3.2; R Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2023).

RESULTS

A total of 469 axSpA patients were included in this longitudinal dataset. This cohort was comprised mostly of males (63%) and r-axSpA patients (75%) (see the online Supplementary Fig 2 for the study flow). Descriptive statistics, including missing data for separate variables, for the 4 patient groups are reported in the online Supplementary Table 1. Our cohort included 117 nr-axSpA patients (67 females and 50 males) and 352 r-axSpA patients (107 females and 245 males). Nr-axSpA patients and females had shorter symptom durations, lower baseline CRP levels, and better baseline spinal mobility than r-axSpA and male patients. Physical function as indicated by the BASFI was worse in r-axSpA males. Although HLA-B27 positivity was more prevalent in r-axSpA, no sex differences were noted in our cohort. Nr-axSpA females comprised as much as 42% of patients with at least 2 comorbidities. Fibromyalgia was more frequently diagnosed among female patients, but no differences were noted between the 2 subtypes. Across the 4 groups, the proportions of participants with peripheral arthritis, psoriasis, uveitis, IBD, and csDMARD use at baseline were similar. Ineffectiveness (ie, the lack or loss of effect of the bDMARD reflecting its performance in real-world settings) was the most common reason for stopping (63%) across all patient groups, followed by adverse events (26%) and other reasons (11%). Nr-axSpA females had the highest proportion (71%) of stoppers due to ineffectiveness (Supplementary Table 2). Patient characteristics according to bDMARD status are found in Supplementary Table 3.

Main effect of subtype and of sex on first-bDMARD discontinuation

Both nr-axSpA patients and females had more first-line bDMARD discontinuations and shorter durations of bDMARD therapy (online Supplementary Table 2). In the bivariable modified Poisson analysis, RRs for first-line bDMARD discontinuance were estimated based on subtype, sex, and each of the predefined categorical and continuous covariates. We found that the nr-axSpA subtype (RR, 1.95; 95% CI, 1.40-2.73; adjusted *P* value < .001) and female sex (RR, 1.82; 95% CI, 1.35-2.44; adjusted *P* value < .001) were associated with a significantly higher risk of bDMARD discontinuation. Having 2 or more comorbidities was likewise linked to an increased discontinuation risk, whereas the risk was lower among patients with a positive HLA-B27 and those using csDMARDs prior to bDMARD initiation (Fig 2A).

On multivariable analysis, the nr-axSpA subtype (RR, 1.80; 95% CI, 1.26-2.58; adjusted *P* = .003, E-value 3.01 [lower bound CI, 1.83]) and female sex (RR, 1.49; 95% CI, 1.08-2.04; adjusted *P* value = .015, E-value 2.34 [lower bound CI, 1.38]) were significantly associated with an increased risk of discontinuation after adjusting for covariates (Fig 2B). The VIF showed the absence of multicollinearity among included variables.

Interaction effect of subtype and sex on first-bDMARD discontinuation

The RERI value pointed to an interaction effect that was 0.49 times greater (95% CI, −0.78 to 1.75; *P* = .450) than the sum of the individual exposures; this was a positive, additive but nonsignificant interaction. The ratio of RRs showed a combined effect of subtype and sex that was 1.05 times greater than their effect product, suggesting a positive, multiplicative interaction, though not significant (95% CI, 0.55-2.03; *P* = .874). Stratified analysis showed that the nr-axSpA subtype had a significantly positive association among females with bDMARD discontinuation (RR, 1.84; 95% CI, 1.16-2.93; adjusted *P* value = .040). Among r-axSpA patients, there was a positive but nonsignificant association between the female sex and discontinuance (RR, 1.46; 95%, 1.00-2.14; adjusted *P* value = .098) (Table 1). With r-axSpA males as the reference, the risk of bDMARD discontinuation in the female nr-axSpA group was 2.7 times higher, indicating a joint effect of the nr-axSpA subtype and female sex (RR, 2.70; 95% CI, 1.73-4.20; adjusted *P* value < .001).

Survival analysis of the first bDMARD

A total of 302 patients (64%) discontinued their first-line bDMARD. BDMARD survival rates showed a decreasing trend over 20 years of follow-up.

BDMARD survival curves according to subtype, sex, and their combination are depicted in Figure 3. Comparing the subtypes, the risk of first-bDMARD termination was significantly increased in nr-axSpA patients compared to those with r-axSpA (HR, 1.65; 95% CI, 1.28-2.12; adjusted *P* value < .001) at any point in time (Fig 3A). The 1-, 5-, 10-, and 15-year drug survival rates were worse in nr-axSpA than in r-axSpA (log-rank test *P* < .001) (Table 2). Between the sexes, the risk of bDMARD discontinuation at any time was 61% higher in females (HR, 1.61; 95% CI, 1.28-2.03; adjusted *P* value < .001) (Fig 3B). BDMARD survival rates (1-, 5-, 10-, and 15-years) were consistently lower in females (log-rank test *P* < .001) (Table 2).

When comparing the 4 patient subgroups, the HR for first bDMARD discontinuation was obtained using the male r-axSpA group as the reference (Fig 3C). The hazard of first bDMARD discontinuation was significantly higher in females than in males in the r-axSpA group (HR, 1.43; 95% CI, 1.08-1.90; adjusted *P* value = .025). Among males, nr-axSpA patients showed a trend towards a higher discontinuation risk versus r-axSpA (HR, 1.38; 95% CI, 0.94-2.03; adjusted *P* value = .101). The discontinuation risk was twice as high in nr-axSpA females compared with r-axSpA males at any time (HR, 2.30; 95% CI, 1.68-3.15; adjusted *P* value < .001). Nr-axSpA females had significantly lower 1-, 5-, 10-, and 15-year first-line bDMARD survival rates (log-rank test *P* < .001) than the other 3 groups (Table 2). The survival curves began to diverge less than 2 years following bDMARD initiation. Median first bDMARD survival rates among the patient groups compared are found in Supplementary Table 4.

We stratified our survival analysis according to the period of bDMARD initiation (Fig 4). Throughout the 3 periods, bDMARD survival was consistently lower in nr-axSpA females, especially in comparison with r-axSpA males (HR, 2.30; 95% CI, 1.68-3.15; *P* < .001). Accounting for the largest difference in bDMARD survival between female nr-axSpA patients and male r-axSpA patients were those who started their first bDMARD before 2009 (HR, 3.52; 95% CI, 1.47-8.44; adjusted *P* value = .014) (Fig 4A). For treatment initiations in the subsequent years, the

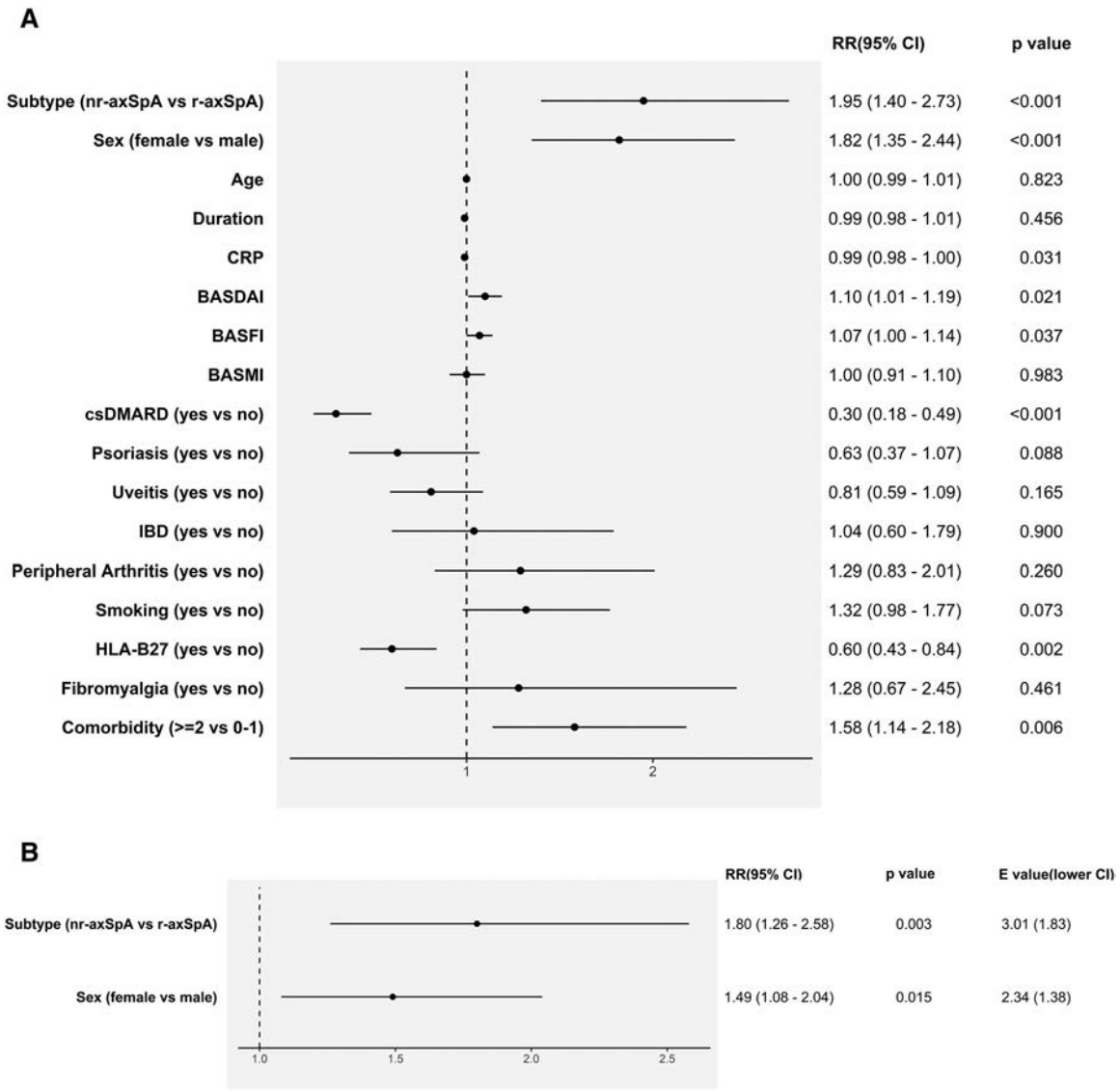


Figure 2. Associations among subtype, sex, covariates, and bDMARD discontinuation. Panels A and B display the crude and adjusted RRs (95% CI), respectively. The vertical line represents a RR of 1. In both panels, *P* values for the main effects of subtype and sex were adjusted for multiplicity using 2 comparisons. (A) The unadjusted *P* values are <.001 for both subtype and sex. (B) The unadjusted *P* values are .0014 for subtype and .015 for sex. bDMARD, biologic disease-modifying antirheumatic drug; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index; CRP, C-reactive protein; csDMARD, conventional synthetic disease-modifying antirheumatic drug; HLA-B27, Human Leukocyte Antigen B27; IBD, inflammatory bowel disease; nr-axSpA, nonradiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis; RR, risk ratio.

risk of bDMARD discontinuation in female nr-axSpA patients decreased in magnitude but remained significantly higher than in male r-axSpA patients between 2009 and April 26, 2016 (HR, 1.85; 95% CI, 1.14-2.98; adjusted *P* value = .037) (Fig 4B) and from April 27, 2016, onwards (HR, 2.63; 95% CI, 1.53-4.53; adjusted *P* value = .001) (Fig 4C).

DISCUSSION

In this study where patients were prospectively followed over 20 years, nr-axSpA patients (versus r-axSpA patients) and female patients (versus male patients) had significantly increased risks of drug discontinuation and lower drug survival rates at all time points (1, 5, 10, and 15 years) after adjusting for varying durations of bDMARD use. We further observed that within the r-axSpA subgroup, females had a 43% higher risk of stopping their first bDMARD compared with males. Our results confirmed observations from previous studies that females with axSpA have shorter bDMARD treatment durations compared to males [29–31]. Data on bDMARD survival in nr-axSpA patients have meanwhile been inconsistent [32–34], but our study revealed that the nr-axSpA subtype also had reduced treatment persistence independent of sex.

The largest difference in bDMARD survival between female nr-axSpA patients and male r-axSpA patients was seen before 2009, during which the recommendations around MRI features highly suggestive of axSpA as well as image acquisition protocols were evolving. This could have led to overdiagnosis of axSpA, especially among females. With the increasing accuracy of nr-axSpA diagnoses over time, discontinuation rates decreased after 2016, although females with nr-axSpA continued to have significantly higher rates of discontinuation. By standardising drug discontinuation rates across all observations using bDMARD treatment duration as an offset term, we verified that female nr-axSpA patients persistently had the lowest bDMARD survival rates throughout the different calendar periods.

Table 1
Interaction between subtype and sex on the risk of first-bDMARD discontinuation

AxSpA subtype	Patient sex				RR (95% CI) of female/ male within the disease subtype stratum
	Male		Female		
	N discontinued/ retained	RR (95% CI)	N discontinued/ retained	RR (95% CI)	
r-axSpA	139/107	1.00 [Reference]	75/31	1.46 (1.00-2.14); $P = .049$ [Female r-axSpA Male r-axSpA]	1.46 (1-2.14); $P = .098$ [Female r-axSpA Male r-axSpA]
nr-axSpA	31/18	1.75 (1.04-2.93) $P = .049$ [Male nr-axSpA Male r-axSpA]	56/11	2.70 (1.73-4.20); $P < .001$ [Female nr-axSpA Male r-axSpA]	1.54 (0.89-2.68); $P = .125$ [Female nr-axSpA Male nr-axSpA]
RR (95% CI) of nr-axSpA/ r-axSpA within the sex stratum		1.75 (1.04-2.93); $P = .098$ [Male nr-axSpA Male r-axSpA]		1.84 (1.16-2.93); $P = .040$ [Female nr-axSpA Female r-axSpA]	

Measure of interaction on the additive scale: RERI (95% CI) = 0.49 (−0.78 to 1.75); $P = .450$.
Measure of interaction on the multiplicative scale: Ratio of RRs (95% CI) = 1.05 (0.55-2.03); $P = .874$.
RRs are adjusted for the following covariates: age, symptom duration, BASDAI, BASFI, BASMI, HLA-B27, CRP, concurrent csDMARD use, psoriasis, uveitis, IBD, peripheral arthritis, smoking, comorbidity count, fibromyalgia.
This table shows adjusted *P* values. For stratified analyses (4 comparisons), unadjusted *P* values are .049 and .125 for female sex (subtype stratum) and .034 and .010 for nr-axSpA subtype (sex stratum). Using male r-axSpA as the reference (3 comparisons), unadjusted *P* values are .049 (female r-axSpA), .034 (male nr-axSpA), and <.001 (female nr-axSpA).
BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; BASMI, Bath Ankylosing Spondylitis Metrology Index; bDMARD, biologic disease-modifying antirheumatic drug; CRP, C-reactive protein; csDMARD, conventional synthetic disease-modifying antirheumatic drug; HLA-B27, Human Leukocyte Antigen B27; HR, hazard ratio; IBD, inflammatory bowel disease; nr-axSpA, nonradiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis; RERI, relative excess risk due to interaction; RR, risk ratio.

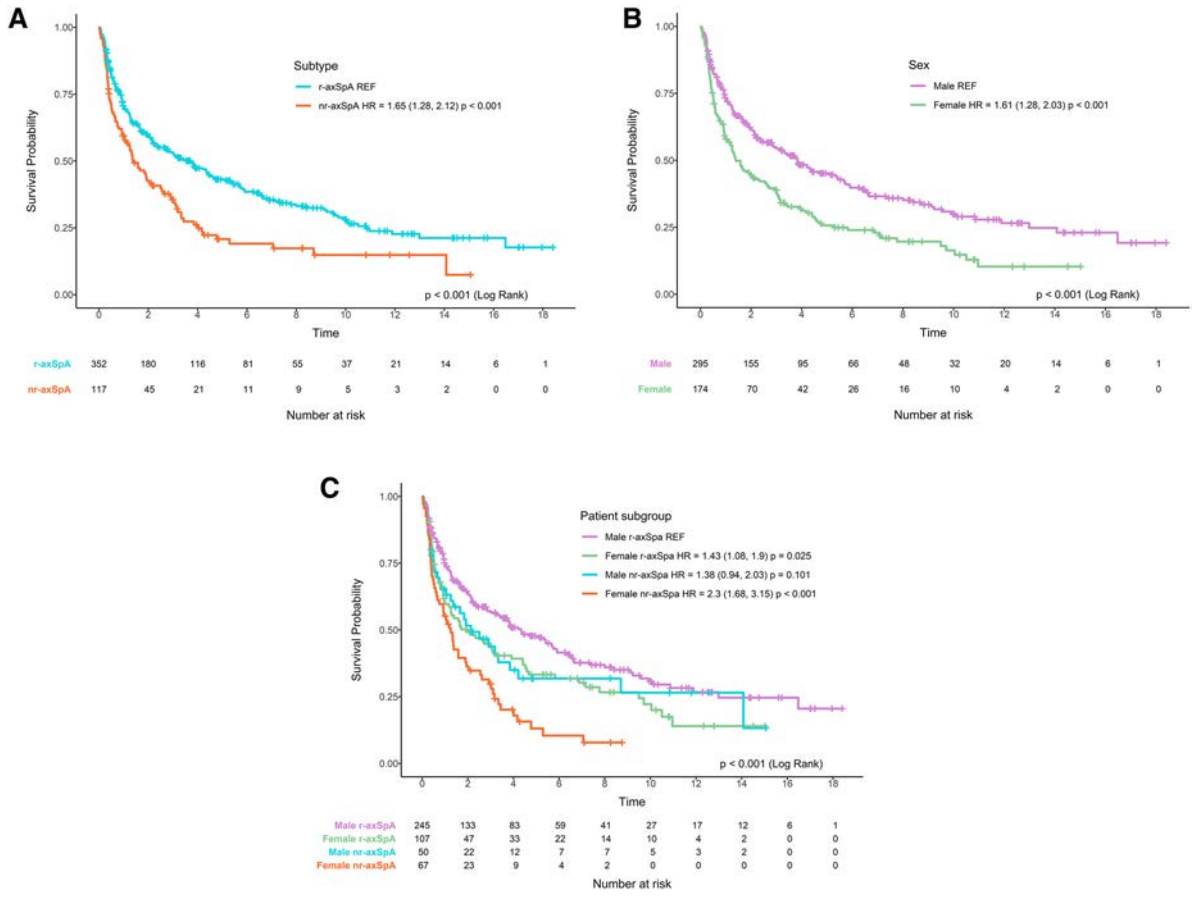


Figure 3. Kaplan–Meier analysis of first-bDMARD discontinuation and log-rank tests according to (A) disease subtype (nr-axSpA versus r-axSpA), (B) patient sex (female versus male), and (C) both subtype and sex. Time (x-axis) is measured in years. bDMARD, biologic disease-modifying antirheumatic drug; HR, hazard ratio; nr-axSpA, nonradiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis; REF, reference group.

Table 2
First-line bDMARD survival rates at 1, 5, 10, and 15 years

Group	n of patients discontinuing bDMARD/Total n	bDMARD survival rates			
		1 year (95% CI)	5 years (95% CI)	10 years (95% CI)	15 years (95% CI)
Subtype					
nr-axSpA	88/117	0.59 (0.51-0.69)	0.21 (0.14-0.31)	0.15 (0.08-0.26)	0.07 (0.02-0.33)
r-axSpA	214/352	0.69 (0.65-0.74)	0.43 (0.38-0.49)	0.28 (0.23-0.35)	0.21 (0.16-0.29)
Sex					
Female	131/174	0.58 (0.51-0.66)	0.26 (0.20-0.34)	0.16 (0.11-0.25)	0.10 (0.05-0.21)
Male	171/295	0.72 (0.67-0.78)	0.45 (0.39-0.52)	0.30 (0.24-0.38)	0.23 (0.17-0.32)
Sex and subtype					
Female nr-axSpA	56/67	0.55 (0.45-0.69)	0.13 (0.06-0.27)	0.08 (0.03-0.22)	0.08 (0.03-0.22)
Male nr-axSpA	32/50	0.65 (0.53-0.80)	0.32 (0.20-0.50)	0.27 (0.15-0.48)	0.13 (0.03-0.60)
Female r-axSpA	75/107	0.60 (0.51-0.70)	0.33 (0.25-0.44)	0.22 (0.14-0.35)	0.14 (0.07-0.28)
Male r-axSpA	139/245	0.74 (0.68-0.80)	0.48 (0.41-0.55)	0.31 (0.24-0.39)	0.25 (0.18-0.34)

Log-rank test *P* values < .001 for all comparisons (see Fig 3).
bDMARD, biologic disease-modifying antirheumatic drug; nr-axSpA, nonradiographic axial spondyloarthritis; r-axSpA, radiographic axial spondyloarthritis.

Similar to our findings from the survival analysis, the nr-axSpA subtype and being female independently conferred a higher risk of drug discontinuation. Moreover, we saw positive interaction trends between the nr-axSpA subtype and female sex. The positive RERI value and ratio of RRs greater than 1 pointed to a small excess risk of first-bDMARD discontinuation in nr-axSpA females, suggesting possible synergy between the 2 exposures at a biological level, although these were not statistically significant. Our results lend support to recent findings from observational studies that female nr-axSpA patients have poorer treatment response and are less likely to achieve inactive disease [6,14,35], likely leading to shorter bDMARD survival. Although the literature on this topic has expanded, our study is the first to employ a factorial design in examining the effect of these variables on the outcome of bDMARD discontinuance. This study allowed us not only to assess the individual effects of sex and subtype on bDMARD survival, but also to determine whether the joint effect of the 2 exposures on the outcome would differ from the main effects. Our relatively small sample size and limited power might have precluded the detection of a

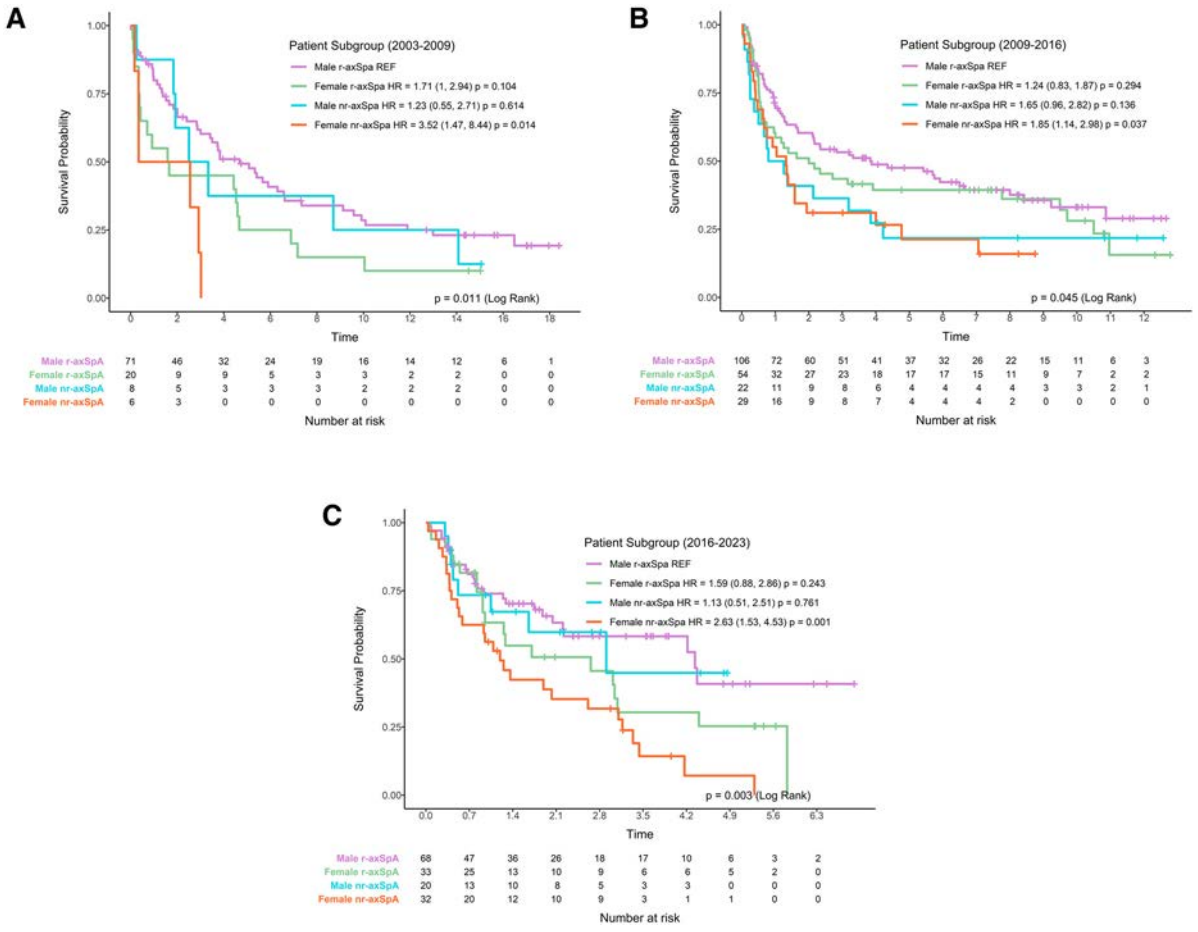


Figure 4. Kaplan–Meier analysis of first-bDMARD discontinuation and log-rank tests. Results reported according to calendar period. (A) 2003-2009, (B) 2009-2016, and (C) 2016-2023. Time (x-axis) is measured in years. bDMARD, biologic disease-modifying antirheumatic drug REF, reference group.

statistically significant interaction. Nonetheless, the direction and magnitude of the RERI and ratio of RRs support our hypothesis for a positive interaction effect.

The reason behind lower bDMARD responsiveness in nr-axSpA females currently remains unclear. The possibility of a higher rate of misdiagnosis among nr-axSpA females has been raised. A study in early axSpA reported females to more frequently fulfil the ASAS clinical arm criteria (50.4% versus 32% in males, $P = .001$) [36]. Moreover, nr-axSpA patients with a positive MRI or elevated baseline CRP showed a trend towards better clinical response to a TNFi versus placebo in some RCTs, compared to those without these objective markers of inflammation [37–39]. In our cohort, as many as 60% to 70% of patients across all 4 subgroups exhibited a response to bDMARD therapy in the initial 6 months, making the likelihood of misdiagnosis relatively low. We could not determine whether the greater risk of discontinuance and high frequency of bDMARD ineffectiveness in nr-axSpA females were associated with meeting the ASAS imaging arm versus clinical arm for nr-axSpA, as this information was not available in all patients.

Although body mass index (BMI) predicted bDMARD survival in some studies, it was not included as a covariate due to its high degree of missingness (56.5%), which could introduce substantial uncertainty or bias in imputed values and affect the robustness of the regression models. The reason behind the large amount of missing BMI data from our database remains to be explored. In our sensitivity analysis, the E-values indicated that the associations between exposures (subtype and sex) and outcome (bDMARD discontinuance) were robust to unmeasured confounding. This indicates that any unmeasured confounder would need to have a very high RR to explain away the associations we observed between the 2 exposures and bDMARD discontinuation. An additional sensitivity analysis excluding mediators was likewise performed to provide a comprehensive perspective. The results yielded estimates consistent with those reported for the direct effects in the main analysis, reinforcing the associations we observed among subtype, sex, and bDMARD discontinuation. However, given the complex interrelationships among covariates, we were unable to perform adjustment for selected confounders without potentially introducing collider bias, highlighting limitations in effect estimation in the context of real-world data. Details of the sensitivity analysis are presented in Supplementary Figure 1.

We chose to assess the effect of subtype and sex on bDMARD discontinuation, with adjustment for baseline parameters, to help identify patient subgroups associated with favourable long-term drug effectiveness and tolerability. As such, time-dependent relationships among disease features, comorbidities, and bDMARD stoppage were not captured. Additionally, we could not account for the effect of antidrug antibodies and sex-dependent biological mechanisms involving genetic associations, hormonal influence, and differences in immune response as these data are not routinely collected in our registry. As comorbidities were analysed as a categorical variable, there was an assumption that individual comorbidities had an equal effect on the outcome. There may be a need to evaluate the relative contribution of each comorbidity in larger studies, especially in the context of the reason behind discontinuation [40]. The diagnosis of fibromyalgia was based on the expert assessment of 2 rheumatologists through patient history and physical examination. As standardised questionnaires were not used, the possibility of missing fibromyalgia in some patients could not be excluded. Being a single-centre study at a tertiary referral centre, our cohort may differ systematically from patients in community practice, which could limit generalizability. However, the similarities between our sample and those in larger studies [7,13,41,42] suggest that our cohort is representative of the broader axSpA population, minimising potential selection bias.

Several methodological strengths distinguish our study. We used multiple imputation to reduce the bias from missing data commonly associated with studies using registry data. Compared to complete case analysis, which yields unbiased estimates only under the restrictive MCAR assumption, multiple imputation provides less biased estimates even when data are not missing at random. The high E-values imply that unmeasured confounding was unlikely to drive our results. The DAG enabled us to incorporate prior knowledge and theoretical understanding of axSpA in covariate selection, ensuring that the regression models were anchored on substantive knowledge rather than purely data-driven methods. The tabular presentation of the additive and multiplicative interactions between subtype and sex improved the transparency and reproducibility of our analyses.

Future studies may benefit from evaluating the impact of the ASAS imaging versus clinical arm on bDMARD survival among female nr-axSpA patients, using the ASAS definition of MRI findings highly suggestive of axSpA [43]. Quantitative methods such as the Spondyloarthritis Research Consortium of Canada scoring system [44] may also be used. These studies can explore whether higher MRI inflammation scores can more accurately identify which female nr-axSpA patients are expected to have better treatment responses. Additionally, future research could explore the relationship among subtype, sex, and specific reasons for bDMARD discontinuation to provide further insights into the observed differences in discontinuation patterns. Similarly, prediction models for axSpA diagnosis and treatment prognosis can be developed for females with nr-axSpA.

In conclusion, we found that discontinuance of the first bDMARD was significantly higher in nr-axSpA than in r-axSpA, in female patients than in males, and in female nr-axSpA patients than all other axSpA patients. The positive additive and multiplicative interactions between subtype and sex, although short of statistical significance, provided some evidence that the 2 exposures act synergistically to shorten the duration of bDMARD therapy in female nr-axSpA patients. Larger studies are needed to reinforce our finding that subtype and sex interact to influence bDMARD discontinuation. Our understanding of sex-related treatment outcomes can be enhanced by including qualitative analyses of Sex, socioeconomic factors, and health-seeking behaviour on the perception of disease control and drug adherence in future studies.

Competing interests

EOS has received honoraria from Johnson and Johnson Philippines, Inc., Novartis Philippines, Pfizer, and Zuellig Pharma. RDI has received grants from Novartis and UCB and consulting fees from Novartis, UCB, Janssen, and Abbvie. NH has received grants from UCB and is part of the advisory board of Abbvie, Novartis, and UCB. PR-R, ESB, TC, ZB, CIAP, and LFD have nothing to disclose.

Acknowledgements

We would like to thank Prof Cynthia Cordero and Dr Adam Zhong for their methodological contributions to this study. An abstract of this study was published as part of the American College of Rheumatology 2024 Convergence.

Contributors

PR-R, ESB, LFD, RDI, and NH designed and conceptualised the study. RDI and NH were involved in patient recruitment. PR-R, TC, ESB, and ZB performed data collection and analysis. LFD, ESB, EOS, CIAP, RDI, and NH made significant contributions to the manuscript's critique and provided valuable intellectual content. PR-R prepared the initial draft, and all authors critically reviewed the manuscript and approved the final version of the manuscript. NH is responsible for the overall content as a guarantor.

Funding

None.

Patient consent for publication

Patients provided informed consent prior to participation in the study.

Ethics approval

This study involved human participants and was approved by the University Health Network Research Ethics Board (REB 23-5976). Participants provided informed consent prior to participation in the study.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

Data may be made available from the senior author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to shorten some sentences to help meet the word count requirement. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Patient and public involvement

Patients and/or the public were not involved in the design, conduct, reporting, or dissemination plans of this research.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.007.

REFERENCES

[1] Bray TJP, Lopes A, Fisher C, Ciurtin C, Sen D, Hall-Craggs MA. Sacroiliac joint ankylosis in young spondyloarthritis patients receiving biologic therapy: observation of serial magnetic resonance imaging scans. Arthritis Rheumatol 2019;71(4):594–8.

[2] van der Heijde D, Baraliakos X, Hermann KA, Landewé RBM, Machado PM, Maksymowych WP, et al. Limited radiographic progression and sustained reductions in MRI inflammation in patients with axial spondyloarthritis: 4-year imaging outcomes from the RAPID-axSpA phase III randomised trial. Ann Rheum Dis 2018;77(5):699–705.

[3] Protopopov M, Poddubnyy D. Radiographic progression in non-radiographic axial spondyloarthritis. Expert Rev Clin Immunol 2018;14(6):525–33.

[4] de Winter J, de Hooge M, van de Sande M, de Jong H, van Hoeven L, de Koning A, et al. Magnetic resonance imaging of the sacroiliac joints indicating sacroiliitis according to the assessment of spondyloarthritis international society definition in healthy individuals, runners, and women with postpartum back pain. Arthritis Rheumatol 2018;70(7):1042–8.

[5] van Onna M, Jurik AG, van der Heijde D, van Tubergen A, Heuft-Dorenbosch L, Landewé R. HLA-B27 and gender independently determine the likelihood of a positive MRI of the sacroiliac joints in patients with early inflammatory back pain: a 2-year MRI follow-up study. Ann Rheum Dis 2011;70(11):1981–5.

[6] Ciurea A, Hebeisen M, Weber U, Tamborrini G, Micheroli R, Wildi LM, et al. In contrast to men, women with nonradiographic axial spondyloarthritis have lower response rates to TNF inhibitors than women with ankylosing spondylitis [abstract]. 2016 ACR/ARHP Annual Meeting: Arthritis Rheumatol; 2016.

[7] Poddubnyy D, Sieper J, Akar S, Muñoz-Fernández S, Haibel H, Hojnik M, et al. Characteristics of patients with axial spondyloarthritis by geographic regions: PROOF multicountry observational study baseline results. Rheumatology (Oxford) 2022;61(8):3299–308.

[8] López-Medina C, Ramiro S, Van Der Heijde D, Sieper J, Dougados M, Molto A. Characteristics and burden of disease in patients with radiographic and non-radiographic axial spondyloarthritis: a comparison by systematic literature review and meta-analysis. RMD Open 2019;5(2):e001108.

[9] Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017;390(10089):73–84.

[10] Wallis D, Haroon N, Ayearst R, Carty A, Inman RD. Ankylosing spondylitis and nonradiographic axial spondyloarthritis: part of a common spectrum or distinct diseases? J Rheumatol 2013;40(12):2038–41.

[11] Landewé R, Braun J, Deodhar A, Dougados M, Maksymowych WP, Mease PJ, et al. Efficacy of certolizumab pegol on signs and symptoms of axial spondyloarthritis including ankylosing spondylitis: 24-week results of a double-blind randomised placebo-controlled Phase 3 study. Ann Rheum Dis 2014;73(1):39–47.

[12] Song IH, Weiß A, Hermann KG, Haibel H, Althoff CE, Poddubnyy D, et al. Similar response rates in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis after 1 year of treatment with etanercept: results from the ESTHER trial. Ann Rheum Dis 2013;72(6):823–5.

[13] Rudwaleit M, Haibel H, Baraliakos X, Listing J, Märker-Hermann E, Zeidler H, et al. The early disease stage in axial spondylarthritis: results from the German Spondyloarthritis Inception Cohort. Arthritis Rheum 2009;60(3):717–27.

[14] Poddubnyy D, Sieper J, Akar S, Muñoz-Fernández S, Haibel H, Diekhoff T, et al. Females with non-radiographic but not with radiographic axial spondyloarthritis are less likely to achieve inactive disease state than males: results from a multicountry prospective observational study. Arthritis Rheumatol 2022;74(suppl 9).

[15] Xie Y, Liu Y, Wu Q. Effect of gender and age on bDMARD efficacy for axial spondyloarthritis patients: a meta-analysis of randomized controlled trials. Rheumatology (Oxford) 2024;63(11):2914–22.

[16] McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003;157(10):940–3.

[17] Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005;162(3):199–200.

[18] Tamhane AR, Westfall AO, Burkholder GA, Cutter GR. Prevalence odds ratio versus prevalence ratio: choice comes with consequences. Stat Med 2016;35(30):5730–5.

[19] Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159(7):702–6.

[20] Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. Int J Epidemiol 2016;45(6):1887–94.

[21] Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. Int J Epidemiol 2012;41(2):514–20.

[22] VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiol Methods 2014;3(1):33–72.

[23] Alli BY. InteractionR: an R package for full reporting of effect modification and interaction. Software Impacts 2021;10:100147.

[24] VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017;167(4):268–74.

[25] Chung WT, Chung KC. The use of the E-value for sensitivity analysis. J Clin Epidemiol 2023;163:92–4.

[26] White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med 2011;30(4):377–99.

[27] Raghunathan TE, Solenberger PW, Van Hoewyk J. IVEware: Imputation and Variance Estimation Software. Ann Arbor, MI: Institute for Social Research, University of Michigan; 2022.
[28] van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1–67.
[29] Glintborg B, Østergaard M, Krogh NS, Tarp U, Manilo N, Loft AG, et al. Clinical response, drug survival and predictors thereof in 432 ankylosing spondylitis patients after switching tumour necrosis factor alpha inhibitor therapy: results from the Danish nationwide DANBIO registry. Ann Rheum Dis 2013;72(7):1149–55.
[30] Rusman T, Ten Wolde S, Euser SM, Van Der Ploeg T, Van Hall O, Van der Horst-Bruinsma I. Gender differences in retention rate of tumor necrosis factor alpha inhibitor treatment in ankylosing spondylitis: a retrospective cohort study in daily practice. Int J Rheum Dis 2018;21:836–42.
[31] Rusman T, van Vollenhoven RF, van der Horst-Bruinsma IE. Gender differences in axial spondyloarthritis: women are not so lucky. Curr Rheumatol Rep 2018;20(6):35.
[32] Ørnbjerg LM, Brahe CH, Askling J, Ciurea A, Mann H, Onen F, et al. Treatment response and drug retention rates in 24 195 biologic-naive patients with axial spondyloarthritis initiating TNFi treatment: routine care data from 12 registries in the EuroSpA collaboration. Ann Rheum Dis 2019;78(11):1536–44.
[33] Glintborg B, Sørensen IJ, Østergaard M, Dreyer L, Mohamoud AA, Krogh NS, et al. Ankylosing spondylitis versus nonradiographic axial spondyloarthritis: comparison of tumor necrosis factor inhibitor effectiveness and effect of HLA-B27 status. An observational cohort study from the nationwide DANBIO registry. J Rheumatol 2017;44(1):59–69.
[34] Michelena X, Zhao SS, Dubash S, Dean LE, Jones GT, Marzo-Ortega H. Similar biologic drug response regardless of radiographic status in axial spondyloarthritis: data from the British Society for Rheumatology Biologics Register in Ankylosing Spondylitis registry. Rheumatology (Oxford) 2021;60(12):5795–800.
[35] Neuenschwander R, Hebeisen M, Micheroli R, Bürki K, Exer P, Niedermann K, et al. Differences between men and women with nonradiographic axial spondyloarthritis: clinical characteristics and treatment effectiveness in a real-life prospective cohort. Arthritis Res Ther 2020;22(1):233.
[36] Tournadre A, Pereira B, Lhoste A, Dubost JJ, Ristori JM, Claudepierre P, et al. Differences between women and men with recent-onset axial spondyloarthritis: results from a prospective multicenter French cohort. Arthritis Care Res (Hoboken) 2013;65(9):1482–9.
[37] Sieper J, van der Heijde D, Dougados M, Mease PJ, Maksymowych WP, Brown MA, et al. Efficacy and safety of adalimumab in patients with non-radiographic axial spondyloarthritis: results of a randomised placebo-controlled trial (ABILITY-1). Ann Rheum Dis 2013;72(6):815–22.
[38] Sieper J, Van Der Heijde D, Dougados M, Maksymowych WP, Scott BB, Boice JA, et al. A randomized, double-blind, placebo-controlled, sixteen-week study of subcutaneous golimumab in patients with active nonradiographic axial spondyloarthritis. Arthritis Rheumatol 2015;67(10):2702–12.
[39] Dougados M, van der Heijde D, Sieper J, Braun J, Maksymowych WP, Citera G, et al. Symptomatic efficacy of etanercept and its effects on objective signs of inflammation in early nonradiographic axial spondyloarthritis: a multicenter, randomized, double-blind, placebo-controlled trial. Arthritis Rheumatol 2014;66(8):2091–102.
[40] Zhao SS, Jones GT, Macfarlane GJ, Hughes DM, Moots RJ, Goodson NJ. Comorbidity and response to TNF inhibitors in axial spondyloarthritis: longitudinal analysis of the BSRBR-AS. Rheumatology (Oxford) 2021;60(9):4158–65.
[41] Baraliakos X, Braun J. Non-radiographic axial spondyloarthritis and ankylosing spondylitis: what are the similarities and differences? RMD Open 2015;1(suppl 1):e000053.
[42] Kiltz U, Baraliakos X, Karakostas P, Igelmann M, Kalthoff L, Klink C, et al. Do patients with non-radiographic axial spondylarthritis differ from patients with ankylosing spondylitis? Arthritis Care Res (Hoboken) 2012;64(9):1415–22.
[43] Lambert RG, Bakker PA, van der Heijde D, Weber U, Rudwaleit M, Hermann KG, et al. Defining active sacroiliitis on MRI for classification of axial spondyloarthritis: update by the ASAS MRI working group. Ann Rheum Dis 2016;75(11):1958–63.
[44] Maksymowych WP, Inman RD, Salonen D, Dhillon SS, Williams M, Stone M, et al. Spondyloarthritis research Consortium of Canada magnetic resonance imaging index for assessment of sacroiliac joint inflammation in ankylosing spondylitis. Arthritis Rheum 2005;53(5):703–9.

Ann Rheum Dis 84 (2025) 594–600

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Systemic lupus erythematosus

Three years is the minimal effective duration of sustained clinical remission which prevents impaired kidney function and damage accrual in lupus nephritis

Mariele Gatto[1], Giulia Frontini[2], Claudia Furlan[3], Marta Calatroni[4,5], Claudio Cruciani[1,6], Francesco Reggiani[4,5], Elisa Bellis[1], Luca Iaccarino[6], Renato Alberto Sinico[4], Gabriella Moroni[4,5], Andrea Doria[6,*]

[1] *Academic Rheumatology Centre, Department of Clinical and Biological Sciences, University of Turin, hospital Mauriziano, Turin, Italy*
[2] *Nephrology and Dialysis Unit, San Paolo Hospital, Milan, Italy*
[3] *Department of Statistical Sciences, University of Padova, Padova, Italy*
[4] *Nephrology and Dialysis Division, Istituti di Ricovero e Cura a Carattere Scientifico, Humanitas Research Hospital, Rozzano, Milan, Italy*
[5] *Department of Biomedical Sciences Humanitas University, Milan, Italy*
[6] *Unit of Rheumatology, Department of Medicine, University of Padova, Padova, Italy*

ARTICLE INFO

Article history:
Received 13 September 2024
Received in revised form 28 December 2024
Accepted 27 January 2025

ABSTRACT

Objectives: To assess the minimum effective duration of remission preventing damage accrual (Systemic Lupus International Collaborating Clinics damage index [SDI]) and impaired kidney function (IKF: estimated glomerular filtration rate of <60 mL/min/1.73 m^2 for at least 3 months) in active lupus nephritis (LN).

Methods: Patients with biopsy-proven LN followed up at least twice yearly were enrolled; clinical variables were collected regularly. Sustained clinical remission (sCR) was defined as estimated glomerular filtration rate of >60 mL/min/1.73 m^2, proteinuria of <0.5 g/24 h and clinical systemic lupus erythematosus disease activity index of 0 for at least 1 year. Log-linear regression and a time-dependent Cox proportional hazard model were used to assess the minimum duration of sCR capable of preventing SDI increase and IKF development.

Results: In total, 293 patients with LN were included (median follow-up: 15.7 [10.4-22.9] years) of whom 84.3% achieved sCR lasting 8.7 (5.4-13.1) years. At last observation, the increase in SDI was higher in patients who never achieved sCR (median: 2 [1-2.5] vs 1 [0-1.5]; $P < .001$). A minimum duration of 3 years of sCR prevented SDI increase (% change = −41.1%; $P = .003$). The analysis on IKF involved only patients with the longest follow-up; 224 patients had ≥10 years of observation. Among them, 50 (22.3%) developed IKF. A minimum duration of 3 years of sCR prevented IKF (hazard ratio = 0.10; $P < .001$). IKF-free survival rate at 10, 20, and 25 years was 87%, 68%, and 40% for patients who never achieved sCR and 99%, 96%, and 91% for patients with at least 3 years of sCR, respectively ($P < .001$).

Conclusions: Three years is the minimum duration of sCR protecting against development of IKF and damage accrual in patients with LN.

*Correspondence to Prof Andrea Doria, Unit of Rheumatology, Department of Medicine, University of Padova, Padova, Italy.
E-mail address: adoria@unipd.it (A. Doria).
MG and GF are equally contributing first authors.
GM and AD are equally contributing senior authors.
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.02.004

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Inadequate renal response leads to overt renal and overall organ damage in patients with lupus nephritis (LN). Remission aids in preserving renal function, yet no firm threshold of remission duration has been established yet.

WHAT THIS STUDY ADDS

- We have shown that 3 years of remission significantly hinder impaired kidney function and Systemic Lupus International Collaborating Clinics damage index accrual, thereby emerging as a safe threshold of remission to be aimed for when treating LN.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- Three years can be pinpointed as the indispensable duration of remission to be pursued when treating patients with LN in clinical practice.

INTRODUCTION

The achievement and maintenance of remission and low disease activity have been widely shown to improve long-term prognosis and protect from irreversible organ damage in systemic lupus erythematosus (SLE) [1,2], whereas recurrent disease flares are among the major drivers of damage and health-related costs [3]. Lupus nephritis (LN) remains one of the most severe manifestations of SLE, which is still leading to terminal renal failure in up to 10% of cases, while jeopardising renal function in virtually all affected patients [4–6]. In fact, the changing paradigm demanding early treatment in LN is mostly advocated by the assumption of LN signifying per se renal damage, regardless of whether this is captured by surrogate laboratory measures such as glomerular filtration rate [7].

Similar to nonrenal SLE, renal flares, and especially nephritic flares, have been since long associated with the poorest renal outcomes [8–10]; hence, the benefits of renal remission were explored in LN as well. We have recently shown that a durable remission involving control of both renal and extrarenal disease activity was associated with a significant decreased likelihood of organ damage accrual and of progression towards terminal renal failure in a multicentric cohort of Caucasian patients with LN and that a remission lasting over 15 years was paired with virtually zero chance of re-flaring [11].

The European recommendations for the management of SLE and LN advise a minimum of 2 years of remission, following at least 3 years of treatment for attempting a safe tapering of immunosuppression in SLE, likely extending this concept to LN [12]; however, no firm data have defined the actual duration of remission hindering overt renal dysfunction and organ damage in LN. In this study, we aimed at assessing the minimum duration of remission (renal and extrarenal) protecting against impaired kidney function (IKF) development and overall organ damage accrual in LN, which would entail practical implications for patient care.

METHODS

Study cohort

This is a retrospective cohort study carried out on an historical cohort of patients with SLE diagnosed between 1970 and 2016 [13]. In this study, we included patients with a biopsy-proven LN classified according to the recent revision of International Society of Nephrology/Renal Pathology Society (2003) [14]. Criteria for inclusion were as follows: (1) SLE classified according to ACR/EULAR criteria; (2) biopsy-proven LN; (3) a follow-up of at least 5 years after the initiation of treatment for LN; and (4) at least 2 evaluations per year. Patients with incomplete records, without kidney biopsy or with pure class II, or with IKF at baseline or requiring renal replacement therapy were excluded from this study.

The study was approved by the Ethics Committee of IRCCS Humanitas Rozzano, Milano, Italy (protocol code NEF0032023). All patients signed an informed consent for the scientific use of their data that was anonymised.

Patient assessment

We considered as baseline the initiation of induction therapy after kidney biopsy. At baseline, at each clinical evaluation, and at last observation data on induction and maintenance therapy, demographics, and renal and overall clinical features were recorded. Disease activity was assessed at each timepoint by SLE disease activity index (SLEDAI)-2K and damage was assessed at baseline and at the end of follow-up through Systemic Lupus International Collaborating Clinics damage index (SDI). IKF occurrence was assessed in patients with at least 10 years of follow-up, in order to include any degree of renal function decline over a considerable period of time from LN diagnosis.

Patient and public involvement statement

Patients or the public were not involved in this retrospective cohort study.

Definition of kidney variables

Sustained clinical remission (sCR) was defined as clinical SLEDAI-2K of 0 (which includes proteinuria <0.5 g/24 h and is irrespective of serology) and estimated glomerular filtration rate (eGFR) of >60 mL/min/1.73 m^2, persisting for at least 1 year with or without glucocorticoids (GCs) and/or immunosuppressive therapy. This definition was validated in a previous study showing its association with protection from organ damage accrual [11]. eGFR was evaluated by modification of diet in renal disease formula and proteinuria measured by benzethonium chloride on the urine collected over 24 hours expressed as in grams per 24 hour. IKF was defined as serum creatinine of >1.0 mg/dL with eGFR of <60 mL/min/1.73 m^2 and inactive urinary sediment, confirmed by at least 3 determinations for at least 3 months. Acute kidney dysfunction was defined as eGFR of <60 mL/min/1.73/m^2 for <3 months, haematuria (urinary red blood cells >20/high-power field [HPF]), and/or erythrocyte casts, and proteinuria of ≥0.5 g/d; renal failure as need of chronic dialysis or eGFR of <15 mL/min/1.73/m^2; nephritic syndrome as increase in serum creatinine of at least 30% over the last value, active urinary sediment (urinary red blood cells >20/HPF and/or erythrocyte casts), and arterial hypertension, with or without increased proteinuria [9]; and arterial hypertension as the mean of 3 consecutive measurements of systolic blood pressure of >140 mm Hg and/or diastolic blood pressure of >90 mm Hg in sitting position.

Statistical analysis

Continuous variables were expressed as median and IQR owing to nonnormal distribution. Mann-Whitney test was used to evaluate changes in median between independent groups of patients, while Wilcoxon signed-rank test was used for paired data. Pearson χ^2 test was used to assess independence between categorical variables.

SDI assessment

We assessed both the percentage of patients accruing damage and the average annual increase in damage, obtained as the overall increase over the duration of the follow-up.

Three log-linear multiple regression models were used to analyse the SDI accrual. Therefore, the score of the SDI accrual (increased by 1) was used as the dependent variable after the log transformation. Moreover, to deal with the paired data structure, the models were controlled by initial SDI.

Model 1 evaluated the effect of the following regressors at their baseline values: AI, chronicity index at kidney biopsy [14], nephritic syndrome, serum creatinine, 24-h proteinuria, arterial hypertension, GC pulses, and hydroxychloroquine (HCQ) use.

The AI and chronicity index reduced the sample size to 202 patients due to the missing values, and because they were never significant in the analyses, they were excluded in the following 2 models to allow a better estimate accuracy.

The steroid dosage was included in model 2 and 3, through the categorisation ≤5 mg/d and >5 mg/d, in line with thresholds accepted for remission in SLE [1]. The sCR duration (in years) was added in model 2 to assess its effect on the SDI accrual. Model 3, with respect to model 2, identified the minimum values of the sCR duration that had a significant protective effect, by categorising the sCR duration in 0, 1, 2, 3, and 4 or more years, with 0 representing the reference category in the models and corresponding to the absence of sCR. Finally, models 1–3 were controlled by ethnicity, sex, age at LN diagnosis, and follow-up duration (in years). Results were given in terms of a % change, that is $(e^{\beta} - 1)100$, where β represent the parameter estimates of the log-linear models.

The variance inflation factor was used to evaluate the multicollinearity among the regressors. The Kolmogorov-Smirnov nonparametric test was used to assess the normality assumption of the models.

IKF assessment

Three Cox proportional hazard models were fit for the IKF occurrence, in a time-dependent framework with respect to the sCR duration, using both the regressors at their baseline values, the steroid dosage and the above-listed 2 sCR duration variables for SDI accrual analyses. In detail, model 4 evaluated the effect of the regressors at their baseline values, model 5 the effect of the sCR duration (in years), and model 6 identified the minimum values of the sCR duration that had a significant protective effect on IKF occurrence. Moreover, models 4–6 were controlled by ethnicity, sex, and age at LN diagnosis.

P values were calculated with (jackknife-corrected) robust standard errors. Proportional hazards assumption was tested globally for each model, through tests based on the scaled Schoenfeld residuals. The proportion of IKF-free patients as a function of the time (years) from LN diagnosis was estimated.

RESULTS

Study cohort

In total, 293 patients with LN were included (median follow-up: 15.7 [10.4-22.9] years) of whom 84.3% achieved sCR lasting 8.7 (5.4-13.1) years after a median 2.0 (1.0-5.0) years from LN onset. Overall, 58 (19.8%) patients presented with acute kidney dysfuction at onset, of whom 45 (77.6%) reached sCR throughout the follow-up period. Baseline demographics and clinical features are reported in Table 1.

Impact of remission duration on the development of chronic damage

At the beginning of the observation, 30 patients (11.8%) had an SDI of >0, while all others did not show any chronic damage. By the end of the observation, 161 patients had their SDI increased. Eighty-eight patients displayed an SDI increase of 1 point, 40 patients of 2 points, 12 patients of 3 points, and 9 patients of 4 points, and 12 patients had it increased of >4 points, up to a maximum of 10 points. The median SDI at the end of follow-up was 1 (0-2). The median difference between the start and end of follow-up was 1 (0-2; $P < .001$.

The sCR duration (years) correlated inversely ($P < .001$) with the increase in SDI (Table 2, model 2), and a duration of 3 years was identified as the shortest period of sCR to significantly protect against SDI increase (model 3; 3 years: $P = .003$; 4 or more years: $P = .001$). Damage increased in 58% vs 86% of patients reaching or not reaching at least 3 years of sCR, respectively (Fig 1A) ($P = .005$). Considering SDI increase, the median

Table 1
Baseline demographics and clinical features of enrolled patients

Variable	All patients (293)
Demographics	
Females, N (%)	256 (87.4)
Caucasians, N (%)	268 (91.5)
Age at SLE diagnosis (y)	25.1 (19.8-33.5)
Age at LN diagnosis (y)	28.7 (23.4-39.2)
Lag time from SLE to LN diagnosis (y)	1.0 (0-6.3)
Clinical parameters	
Serum creatinine (mg/dL)	0.8 (0.7-1.2)
eGFR (mL/min/1.73 m^2)	82.0 (55.6-98.9)
Acute kidney dysfunction, N (%)	58 (19.8)
Proteinuria (g/d)	3.5 (2.0-5.4)
Arterial hypertension, N (%)	125 (43.0)
SDI equal to 0, N (%)	224 (88.2)
LN characteristics	
Class V/III, IV, or mixed, N (%)	52 (17.8)/241 (82.2)
Activity index	6 (3-9)
Chronicity index	1 (0-3)
Clinical renal syndromes at baseline, n (%)	
Asymptomatic urinary abnormalities	112 (38.2)
Nephrotic syndrome	123 (42.0)
Nephritic syndrome	54 (18.4)
Rapidly progressive renal failure	4 (1.4)
Treatment, N (%)	
Methylprednisolone pulses induction	231 (79.4)
Immunosuppressive therapy induction	252 (87.8)
CYC/AZA/MMF/others	155 (61.5)/22 (8.7)/ 63 (25.0)/12 (4.8)
Hydroxychloroquine	88 (33.6)

AZA, azathioprine; CYC, cyclophosphamide; eGFR, estimated glomerular filtration rate; LN, lupus nephritis; MMF, mycophenolate mofetil; SDI, Systemic Lupus International Collaborating Clinics damage index; SLE, systemic lupus erythematosus.
Continuous variables expressed as median (IQR).

Table 2
Results of the log-linear multiple regression models for SDI accrual, with baseline SDI as a regressor for the paired data structure

Regressors	Model 1			Model 2			Model 3		
	% Change	95% CI	*P*	% Change	95% CI	*P*	% Change	95% CI	*P*
Baseline SDI	41.1***	14.7, 73.5	.001	48.1****	25.5, 74.7	<.001	52.6****	29.0, 80.5	<.001
Baseline values									
Activity index	0.5	−1.6, 2.7	.624						
Chronicity index	2.2	−2.1, 6.7	.315						
Nephritic syndrome	47.6***	11.2, 95.9	.007	39.2***	10.8, 74.9	.005	45.1***	14.1, 84.4	.003
Serum creatinine	−11.3	−23.4, 2.8	.112	−10.2	−21.5, 2.7	.117	−10.5	−22, 2.5	.109
24-h proteinuria	1.3	−1.0, 3.6	.274	1.0	−0.9, 3.0	.308	1.7	−0.3, 3.8	.105
Arterial hypertension	31.5***	9.5, 57.7	.004	18.9**	2.9, 37.4	.019	21.0**	4.5, 40.1	.011
GC pulses	−18.7	−36.8, 4.6	.107	−20.1**	−32.6, −5.2	.010	−19.2**	−32.1, −3.8	.017
HCQ use	−20.1**	−36.0, −0.2	.048	−17.2**	−29.4, −2.8	.021	−13.9*	−27.1, 1.6	.076
Steroid dosage: >5 vs ≤5 (mg/d)				−2.6	−15.5, 12.3	.716	−9.4	−22.1, 5.3	.197
sCR duration (y)				−2.5****	−3.7, −1.4	<.001			
sCR duration (y)									
1 vs 0							−24.8	−54.2, 23.7	.260
2 vs 0							−20.4	−44.5, 14.3	.216
3 vs 0							−41.1***	−58.4, −16.4	.003
4 or more vs 0							−33.5****	−46.9, −16.7	<.001
Intercept	49.1	−7.5, 140.4	.100	58.2**	7.3, 133.2	.021	124.7****	45.4, 247.1	<.001

GC, glucocorticoid; HCQ, hydroxychloroquine; LN, lupus nephritis; sCR, sustained clinical remission; SDI, Systemic Lupus International Collaborating Clinics damage index.
Model 1 assessed the effect of the listed regressors at their baseline values, model 2 of the sCR duration, and model 3 of the minimum protective sCR duration. The models were controlled by confounding variables (follow-up duration, ethnicity, sex, and age at LN diagnosis [y]).
Normality assumption was met ($P > .05$ in each model). Multicollinearity was excluded: variance inflation factor < 3 in each model.
Significance level: *$P < .10$; **$P < .05$; ***$P < .01$; ****$P < .001$.

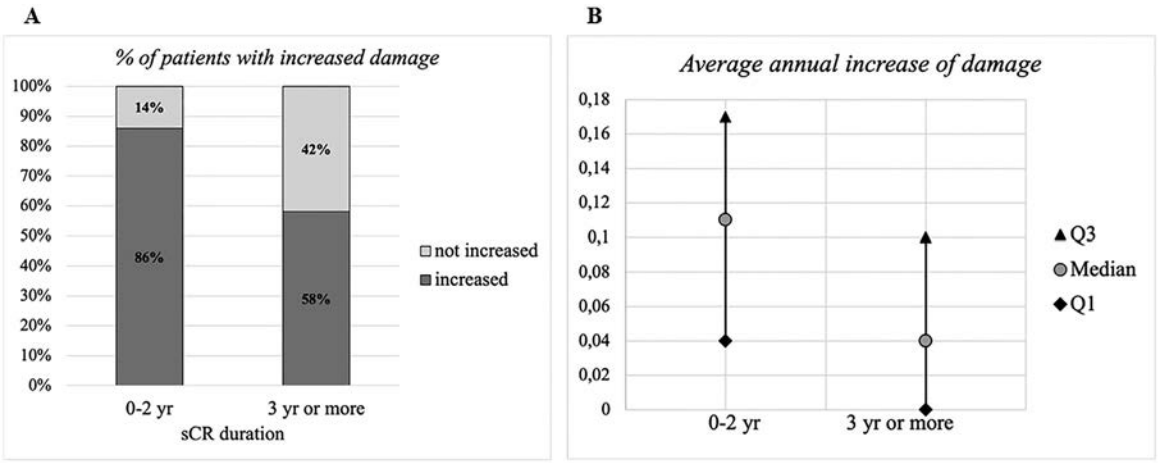


Figure 1. A, percentage of patients with increased damage, with sCR duration of <3 or ≥3 years. B, median, first (Q_1) and third (Q_3) quartiles of the average annual increase of the damage with sCR duration <3 or ≥3 years. sCR, sustained complete remission.

average annual increase of damage was 0.04 points vs 0.11 points for sCR of ≥3 years or <3 years, respectively ($P < .001$) (Fig 1B).

Regarding the significant effects of the regressors at their baseline values (model 3), SDI accrual was 45.1% higher if nephritic syndrome was present at onset ($P = .003$), 21.0% higher in case of arterial hypertension ($P = .011$), and 19.2% lower in patients using GC pulses ($P = .017$) vs those who did not.

Impact of remission duration on the development of IKF

In our cohort, 224 patients had a minimum of 10 years of follow-up, of whom 30 never achieved sCR. Among the remaining 194 patients, 191 had a remission duration of at least 2 years, 179 of at least 3 years, and 174 of at least 4 years. Overall, 50 of the 224 patients developed IKF. The clinical and demographical features of this patient subgroup are depicted in Supplementary Table.

The minimum duration of sCR to be a significant protective factor for IKF occurrence (Table 3; model 6) was 3 years ($P < .001$; 4 or more years: $P < .001$). In detail, compared with patients who never achieved remission, the probability of IKF decreased by 90% (hazard ratio [HR] = 0.10; [HR−1] 100 = −90) with 3 years of sCR, and by 91% (HR = 0.09; [HR−1]100 = −91] with 4 or more years.

The IKF-free survival at 10, 20, and 25 years was 87%, 68%, and 40% for patients who never achieved sCR and 99%, 96%, and 91% for patients with at least 3 years of sCR, respectively ($P < .001$). Figure 2 shows the curves of the estimated proportion of IKF-free patients separated by remission duration.

DISCUSSION

In this retrospective study on a multicentric SLE cohort, we assessed the minimum duration of remission capable of protecting against overt impaired renal dysfunction and overall organ

Table 3
Results of the Cox models for IKF with sCR duration as time-dependent regressor

Regressors	Model 4			Model 5			Model 6		
	HR	95% CI	*P*	HR	95% CI	*P*	HR	95% CI	*P*
Baseline values									
Activity index	1.06	0.94, 1.18	.352						
Chronicity index	1.17*	0.97, 1.40	.095						
Nephritic syndrome	2.76	0.68, 11.18	.156	1.06	0.34, 3.27	.917	2.24	0.73, 6.85	.157
Serum creatinine	0.88	0.39, 1.96	.747	1.90	0.87, 4.16	.109	1.33	0.64, 2.76	.442
24-h proteinuria	0.94	0.82, 1.07	.318	0.90	0.80, 1.02	.109	0.94	0.82, 1.08	.364
Arterial hypertension	2.64**	1.18, 5.90	.018	1.66	0.70, 3.93	.247	1.93	0.87, 4.29	.106
GC pulses	0.59	0.25, 1.38	.222	0.46*	0.19, 1.08	.075	0.65	0.29, 1.45	.289
HCQ use	0.51	0.05, 4.77	.556	0.91	0.25, 3.30	.887	0.75	0.16, 3.46	.714
Steroid usage: >5 vs ≤5 (mg/d)				1.80	0.82, 3.98	.144	1.31	0.65, 2.64	.443
sCR duration (y)				0.78****	0.72, 0.85	<.001			
sCR duration (y)									
1 vs 0							0.56	0.06, 5.18	.612
2 vs 0							0.91	0.33, 2.49	.851
3 vs 0							0.10****	0.04, 0.29	<.001
4 or more vs 0							0.09****	0.03, 0.27	<.001

GC, glucocorticoid; HCQ, hydroxychloroquine; HR, hazard ratio; IKF, impaired kidney function; LN, lupus nephritis; sCR, sustained clinical remission; SDI, Systemic Lupus International Collaborating Clinics damage index.
Model 4 assessed the effect of the listed regressors at their baseline values, model 5 of the sCR duration, and model 6 of the minimum protective sCR duration. The models are controlled by confounding variables (ethnicity, sex, and age at LN diagnosis). Proportional hazard assumption was globally met ($P > .05$) in each model.
Significance level: *$P < .10$; **$P < .05$; ***$P < .01$; ****$P < .001$.

damage in patients with SLE and active LN. We found that at least 3 years of remission safely separate patients who will or will not develop IKF and significant SDI accrual, regardless of the underlying treatment.

The issue of remission in LN has been long addressed in the last years, paralleling the change in the treatment paradigm, which envisions an early intensified treatment as the optimised approach to LN, leading to achievement and especially maintenance of a durable renal response over time, in turn helping to preserve long-term renal function [7,15,16]. Previous cohort studies and expert recommendations advise a duration of clinical remission of at least 2 years to significantly delay damage progression in SLE [12,17,18]; although this concept could be inferred to severe manifestations, no such information is specifically available for LN.

While, intuitively, the longer the time spent in remission, the better the long-term outcome, no firm data have set an ideal threshold for remission to be aimed for in order to minimise progression towards renal failure. Additionally, the impending concept of LN highlights that all affected patients are burdened with renal dysfunction because of the silent nephron loss accelerating at each LN flare [19,20], therefore urging to prevent any identifiable degree of decline in renal function, before overt renal failure occurs. Consequently, and because renal involvement is inserted into a more complex picture in SLE patients, we also considered clinical SLEDAI-2K, in addition to eGFR in our definition of remission [11] and IKF over terminal renal failure as the study end point.

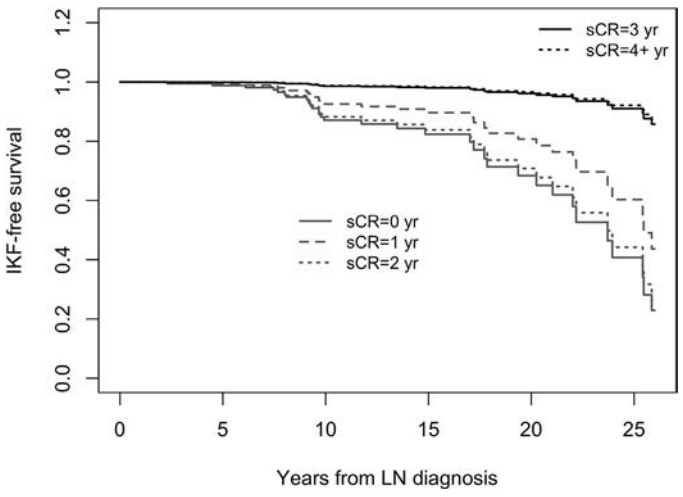


Figure 2. Proportion of IKF-free patients as a function of the time (years) from LN diagnosis, estimated by model 6, for each value of the sCR duration. Other regressors were fixed at their average values. IKF, impared kidney function; LN, lupus nephritis; sCR, sustained clinical remission.

Our observations suggest that pursuing at least 3 years of clinical remission encompassing both the renal and extrarenal domains could significantly improve prognosis in LN patients by avoiding IKF in most of them. Interestingly, no difference in renal function decline was apparent until up to 10 years from LN onset, while the difference arises to be significant over that timespan (Fig 2). This suggests, on the one hand, that eGFR itself is at best an average surrogate for renal function, not accurately mirroring the true nephron loss occurring since LN onset, and, on the other hand, that an adequate remission duration is required to provide a beneficial effect in the long term.

Among independent predictors of SDI increase and IKF development, uncontrolled arterial hypertension plays a major role, in keeping with previous data showing its significant effect on renal outcome in LN through the exploitation of nonimmune mechanisms of renal deterioration involving hyperfiltration and angiosclerosis in the first place [13,20–22], underscoring the need for prevention of comorbidities to halt progression towards overall and kidney-limited damage [23].

Importantly, higher dosages of oral GC at the time of sCR did not increase the chance of damage or IKF in our cohort, suggesting that control of disease activity plays a major role in avoiding damage progression during the acute stages of disease, consistently with former data from the Padova lupus cohort [1]. Interestingly, use of GC pulses was an independent negative predictor of SDI increase in our cohort, which reinforces the recommendation of preferring pulses over oral GCs in active stages

of disease [18]. This may be due to the differential biological effects of GC pulses, that is, exploiting more rapid nongenomic effects [24] together with the higher likelihood of a quicker tapering associated with initial intravenous administration [25,26].

HCQ intake is recommended in all patients with SLE including those with active LN and is known to foster a number of beneficial effects in SLE, accompanied by a very good safety profile [18,27,28]; however, HCQ intake in our cohort did not immediately discriminate patients developing or not IKF, when tested against distinct remission thresholds in the statistical model. This might have been affected by the relatively small amount of patients receiving antimalarials at disease onset, which may be due to the historical nature of the cohort covering a remarkable timespan [21], up to when use of antimalarials was not usual in LN. Nevertheless, the protection conferred by HCQ appears to play a relevant role in the long term, which may require a longer follow-up and different study goals to appear; this is also suggested by HCQ intake to indeed significantly prevent damage in our cohort (Table 2, model 2).

It is worth noting that proteinuria did not emerge as a reliable indicator of IKF or damage in our cohort, which is in keeping with its reliability being questioned as a long-standing biomarker in LN [29,30]. Similarly, histologic lesions at index biopsy did not inform on further evolution, in line with previous findings from different cohorts [21,31]. This observation raises a burning issue in clinical practice, that is, the need for reliable and early measurable biomarkers that would allow monitoring of renal disease and anticipation of response or flares from the early stages of disease. In this regard, emerging soluble urinary biomarkers seem promising, especially encompassing macrophage and neutrophil-related molecules, which undergo early changes upon treatment initiation for active LN, thereby predicting renal response and outperforming proteinuria [32]. How these variations actually associate with a durable remission remains to be explored.

In conclusion, our data show for the first time that a sustained remission involving both renal and extrarenal domains lasting at least 3 years is clinically meaningful in protecting patients with LN from IKF development and SDI increase. Limitations of our study encompass the issue of ethnicity, which prevents generalisation to non-Caucasian cohorts, a percentage of missing data including comorbidities and use of nonimmunosuppressants drugs relevant in LN (eg, renin-angiotensin-aldosterone system) and the joint consideration of continuous and cumulative remission, whose differentiation would require a larger cohort and a prospective design. Altogether, our observations can add a useful piece of information to practical management of patients with LN in that they help setting an objective threshold of remission duration to be pursued when treating LN in clinical practice.

Competing interests

All authors have nothing to declare.

Contributors

GM, AD, MG, and GF contributed to conceptualisation of the study. MG, GF, CF, GM, MC, CC, FR, EB, LI, and RAS curated the data. CF, MG, and GF performed formal analysis. MG, GF, GM, and AD contributed to investigation. CF, MG, GF, GM, and AD contributed to the methodology, supervised the study, and validated the data. MG, GF, CF, MC, CC, FR, EB, LI, RAS, GM, and AD visualised the study. MG, GF, and CF wrote the original draft. All authors reviewed, edited, and approved the final version of the manuscript. MG, GF, GM, and AD are the guarantors of the study.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Patient consent for publication

All patients signed an informed consent for the scientific use of their data that was anonymised.

Ethics approval

The study was approved by the Ethics Committee of Istituti di Ricovero e Cura a Carattere Scientifico Humanitas Rozzano, Milano, Italy (protocol code NEF0032023).

Provenance and peer review

Not commissioned, externally peer reviewed.

Data availability statement

Data are available upon reasonable request.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.02.004.

REFERENCES

[1] Zen M, Iaccarino L, Gatto M, Bettio S, Saccon F, Ghirardello A, et al. The effect of different durations of remission on damage accrual: results from a prospective monocentric cohort of Caucasian patients. Ann Rheum Dis 2017;76:562–5.
[2] Golder V, Kandane-Rathnayake R, Li N, Louthrenoo W, Chen YH, Cho J, et al. Association of sustained lupus low disease activity state with improved outcomes in systemic lupus erythematosus: a multinational prospective cohort study. Lancet Rheumatol 2024;6:e528–36.
[3] Doria A, Amoura Z, Cervera R, Khamastha MA, Schneider M, Richter J, et al. Annual direct medical cost of active systemic lupus erythematosus in five European countries. Ann Rheum Dis 2014;73:154–60.
[4] Tektonidou MG, Dasgupta A, Ward MM. Risk of end-stage renal disease in patients with lupus nephritis, 1971-2015: a systematic review and Bayesian meta-analysis. Arthritis Rheumatol 2016;68:1432–41.
[5] Hanly JG, O'Keeffe AG, Su L, Urowitz MB, Romero-Diaz J, Gordon C, et al. The frequency and outcome of lupus nephritis: results from an international inception cohort study. Rheumatology (Oxford) 2016;55:252–62.
[6] Moroni G, Gatto M, Tamborini F, Quaglini S, Radice F, Saccon F, et al. Lack of EULAR/ERA-EDTA response at 1 year predicts poor long-term renal outcome in patients with lupus nephritis. Ann Rheum Dis 2020;79:1077–83.
[7] Lichtnekert J, Anders H-J, Lech M. Lupus nephritis: current perspectives and moving forward. J Inflamm Res 2022;15:6533–52.
[8] Yap DYH, Tang CSO, Ma MKM, Lam MF, Chan TM. Survival analysis and causes of mortality in patients with lupus nephritis. Nephrol Dial Transplant 2012;27:3248–54.

[9] Moroni G, Quaglini S, Maccario M, Banfi G, Ponticelli C. "Nephritic flares" are predictors of bad long-term renal outcome in lupus nephritis. Kidney Int 1996;50:2047–53.
[10] Parikh SV, Nagaraja HN, Hebert L, Rovin BH. Renal flare as a predictor of incident and progressive CKD in patients with lupus nephritis. Clin J Am Soc Nephrol 2014;9:279–84.
[11] Gatto M, Frontini G, Calatroni M, Reggiani F, Depascale R, Cruciani C, et al. Effect of sustained clinical remission on the risk of lupus flares and impaired kidney function in patients with lupus nephritis. Kidney Int Rep 2024;9:1047–56.
[12] Fanouriakis A, Kostopoulou M, Cheema K, Anders HJ, Aringer M, Bajema I, et al. 2019 Update of the Joint European League Against Rheumatism and European Renal Association–European Dialysis and Transplant Association (EULAR/ERA–EDTA) recommendations for the management of lupus nephritis. Ann Rheum Dis 2020;79:713–23.
[13] Moroni G, Vercelloni PG, Quaglini S, Gatto M, Gianfreda D, Sacchi L, et al. Changing patterns in clinical–histological presentation and renal outcome over the last five decades in a cohort of 499 patients with lupus nephritis. Ann Rheum Dis 2018;77:1318–25.
[14] Bajema IM, Wilhelmus S, Alpers CE, Bruijn JA, Colvin RB, Cook HT, et al. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. Kidney Int 2018;93:789–96.
[15] Parodis I, Houssiau FA. From sequential to combination and personalised therapy in lupus nephritis: moving towards a paradigm shift? Ann Rheum Dis 2022;81:15–9.
[16] Gatto M, Zen M, Cruciani C, Iaccarino L, Doria A. Navigating the landscape of SLE treatment: an expert viewpoint on the rationality and limitations of early biologic intervention. Autoimmun Rev 2024;23:103612.
[17] Zen M, Saccon F, Gatto M, Montesso G, Larosa M, Benvenuti F, et al. Prevalence and predictors of flare after immunosuppressant discontinuation in patients with systemic lupus erythematosus in remission. Rheumatology (Oxford) 2020;59:1591–8.
[18] Fanouriakis A, Kostopoulou M, Andersen J, Aringer M, Arnaud L, Bae SC, et al. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. Ann Rheum Dis 2024;83:15–29.
[19] Luyckx VA, Rule AD, Tuttle KR, Delanaye P, Liapis H, Gandjour A, et al. Nephron overload as a therapeutic target to maximize kidney lifespan. Nat Rev Nephrol 2022;18:171–83.
[20] Anders H-J, Rovin B. A pathophysiology-based approach to the diagnosis and treatment of lupus nephritis. Kidney Int 2016;90:493–501.
[21] Gatto M, Radice F, Saccon F, Calatroni M, Frontini G, Trezzi B, et al. Clinical and histological findings at second but not at first kidney biopsy predict end-stage kidney disease in a large multicentric cohort of patients with active lupus nephritis. Lupus Sci Med 2022;9:e000689.
[22] Rong R, Wen Q, Wang Y, Zhou Q, Qiu Y, Lu M, et al. Prognostic significance of hypertension at the onset of lupus nephritis in Chinese patients: prevalence and clinical outcomes. J Hum Hypertens 2022;36:153–62.
[23] Gatto M, Zen M, Iaccarino L, Doria A. New therapeutic strategies in systemic lupus erythematosus management. Nat Rev Rheumatol 2019;15:30–48.
[24] Stahn C, Buttgereit F. Genomic and nongenomic effects of glucocorticoids. Nat Clin Pract Rheumatol 2008;4:525–33.
[25] Ruiz-Irastorza G. Prednisone in systemic lupus erythematosus: taper quickly, withdraw slowly. Rheumatology (Oxford) 2021;60:5489–90.
[26] Ruiz-Irastorza G, Ruiz-Estevez B, Lazaro E, Ruiz-Arruza I, Duffau P, Martin-Cascon M, et al. Prolonged remission in SLE is possible by using reduced doses of prednisone: an observational study from the Lupus-Cruces and Lupus-Bordeaux inception cohorts. Autoimmun Rev 2019;18:102359.
[27] Ruiz-Irastorza G, Martín-Iglesias D, Soto-Peleteiro A. Update on antimalarials and systemic lupus erythematosus. Curr Opin Rheumatol 2020;32:572–82.
[28] Jorge A, McCormick N, Lu N, Zheng Y, Esdaile J, De Vera M, et al. Hydroxychloroquine and mortality among patients with systemic lupus erythematosus in the general population. Arthritis Care Res (Hoboken) 2021;73:1219–23.
[29] Zickert A, Sundelin B, Svenungsson E, Gunnarsson I. Role of early repeated renal biopsies in lupus nephritis. Lupus Sci Med 2014;1:e000018.
[30] Carlucci PM, Li J, Fava A, Deonaraine KK, Wofsy D, James JA, et al. High incidence of proliferative and membranous nephritis in SLE patients with low proteinuria in the Accelerating Medicines Partnership. Rheumatology (Oxford) 2022;61:4335–43.
[31] Parodis I, Adamichou C, Aydin S, Gomez A, Demoulin N, Weinmann-Menke J, et al. Per-protocol repeat kidney biopsy portends relapse and long-term outcome in incident cases of proliferative lupus nephritis. Rheumatology (Oxford) 2020;59:3424–34.
[32] Fava A, Buyon J, Magder L, Hodgin J, Rosenberg A, Demeke DS, et al. Urine proteomic signatures of histological class, activity, chronicity, and treatment response in lupus nephritis. JCI Insight 2024;9:e172569. doi: 10.1172/jci.insight.172569.

Ann Rheum Dis 84 (2025) 601–608

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Sjögren's syndrome

Increasing the number of minor salivary glands from patients with Sjögren's disease improves the diagnostic and measurement precision of the histological focus score

Konstantinos Tryposkiadis[1,2], Saba Nayar[3], Valentina Pucino[2,3], Charlotte G. Smith[3], Rachel M. Brown[4], Timothy Bates[4], Simon J. Bowman[2,3], Alice Sitch[1,2], Malcolm Price[1,2,5], Francesca Barone[2,3,6], Jon Deeks[1,2], Benjamin A. Fisher[2,3,*]

[1] *Test Evaluation Research Group, Department of Applied Health Sciences, University of Birmingham, Birmingham, UK*
[2] *NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK*
[3] *Rheumatology Research Group, School of Infection, Inflammation and Immunology, University of Birmingham, Birmingham, UK*
[4] *Department of Histopathology, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK*
[5] *Department of Public Health, Canadian University of Dubai, Dubai*
[6] *Candel Therapeutics, Needham, MA, USA*

ARTICLE INFO

Article history:
Received 7 October 2024
Received in revised form 2 December 2024
Accepted 27 December 2024

ABSTRACT

Objectives: Minor salivary gland (MSG) biopsy has an important role in Sjögren's disease diagnosis and research. MSGs show within-patient variation in number of lymphocytic foci per unit area, but the optimal number of MSGs required to balance reproducibility and clinical acceptability has not been determined.

Methods: Monte Carlo simulations were performed to investigate impact of MSG number on (i) diagnosis based on focus score (FS) ≥1; (ii) reproducibility, defined as the extent to which 2 FS measurements obtained from 2 within-patient biopsies are the same, assuming no systematic differences have occurred in between biopsies; and (iii) smallest sample size required to detect a clinically meaningful difference in FS. Data simulation was repeated for different MSG numbers (range, 2-7).

Results: Higher reproducibility was noted for every unit increase in MSG number, with the median absolute difference between 2 within-patient FS measurements decreasing from 1.05 (SD = 0.25) with 2 glands to 0.52 (SD = 0.12) with 7 glands. MSG number influenced the probability of a simulated patient receiving a FS ≥1, increasing from a median of 0.67 with 2 glands to 0.77 with ≥5 glands. MSG number influenced clinical trial sample sizes. For example, 80% statistical power to detect a 40% FS reduction required a sample size per group of 62 with 2 glands and 25 with 7 glands.

Conclusions: For a diagnostic threshold of FS ≥1, a minimum of 5 glands should ideally be targeted. For continuous FS values, a larger number of MSGs (eg, 6) will increase reproducibility further and reduce clinical trial sample size requirements.

*Correspondence to Prof. Benjamin A. Fisher, Rheumatology Research Group, School of Infection, Inflammation and Immunology, College of Medicine and Health, University of Birmingham, Birmingham, UK.

E-mail address: B.Fisher@bham.ac.uk (B.A. Fisher).

Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.01.038

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Salivary gland histopathology has a pivotal role in the diagnosis and classification of Sjögren's disease (SjD) and is widely used as a secondary or exploratory outcome measure in clinical trials.
- The focus score (FS) is the key metric, which is a count of the number of lymphocytic foci per unit area.
- There is variation in the number of foci between individual minor salivary glands (MSGs) from the same patient, but there are no data on the optimal number of salivary glands for diagnostic and measurement accuracy.
- Recommendations on histopathology for SjD have advised research on the optimal number of MSGs to ensure a reproducible result.

WHAT THIS STUDY ADDS

- An increase in MSG number is associated with increased FS reproducibility and deceased sample size requirements.
- For diagnosis, 5 MSGs in a biopsy (or 4 if large) is optimal for FS evaluation.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- We anticipate that our data will contribute to more accurate diagnosis and classification of SjD.
- Our data will inform research seeking to correlate salivary gland pathology with peripheral biomarkers or clinical outcomes or with response to treatment.
- Alternative tissue trial outcome measures should continue to be explored.

INTRODUCTION

Sjögren's disease (SjD) is characterised by autoimmune inflammation of exocrine glands resulting in severe dryness of the eyes and mouth, fatigue, and systemic manifestations in a proportion of patients. Lymphoma, often arising within the salivary glands, is a late complication in 5% of patients [1,2]. SjD affects 14 times as many women as men and has a peak age of onset of around 50 years [3]. Distinguishing SjD from other causes of symptomatic dryness is a diagnostic challenge.

Salivary gland biopsy has a pivotal role in the diagnosis of SjD, with a positive biopsy contributing high weight within the most recent 2016 American College of Rheumatology/European Alliance of Associations for Rheumatology (ACR/EULAR) classification criteria and being essential for classification in patients who are anti-Ro/Sjögren's syndrome antigen A antibody negative. Inflammation within the salivary glands appears as lymphocytic foci, also known as tertiary lymphoid structures, with progressive degrees of organisation including germinal centre formation reminiscent of that seen in secondary lymphoid organs. The presence of isolated lymphocytic foci is not in itself diagnostic of SjD, meaning that histological assessment includes a 2-step process. The first is identification of a pattern of focal lymphocytic sialadenitis, ie, the presence of dense aggregates (foci) of ≥50 mononuclear cells in a periductal location [4–6]. The second is enumeration of the number of foci over the background glandular area with the established diagnostic threshold being a focus score (FS) of ≥1 foci per 4 mm^2.

In addition to its diagnostic value, a high FS has been associated with increased risk for future lymphoma development [7–9]. Salivary gland histopathology measures, including the FS, have also been explored as SjD clinical trial outcomes [5,6,10–13].

Despite its importance in the assessment of SjD, some studies have highlighted variation between laboratories in histological reporting [14], although experienced independent observers with consistent methodology can achieve similar results [15]. An additional challenge to obtaining a representative FS is stochastic variation in the number of foci observed in tissue sections between individual glands within the same patient. Increasing the number of minor salivary glands (MSGs) obtained may overcome the effect of this between-gland within-patient variability. Recommendations published in 2017 suggested obtaining 4 to 6 glands, but this number was not data-driven [6] and in many centres, fewer glands are obtained. The agenda for future work in those Recommendations included establishing the optimal tissue requirements to achieve a reproducible result.

METHODS

Monte Carlo simulations were performed to investigate impact of MSG number on (i) diagnosis based on FS ≥1, (ii) reproducibility of FS, and (iii) the smallest sample size required to detect a clinically meaningful difference in the FS. The simulations were based on the following scenarios:

- For (i), patients with SjD underwent MSG biopsy, with a FS measurement obtained from each biopsy. The probability of FS ≥1 was expressed as the number of patients with a FS ≥1 over the total number of patients.
- For (ii), patients with SjD underwent 2 consecutive MSG biopsies. The reproducibility was expressed as the average absolute difference between the 2 FS measurements obtained from each patient's biopsies.
- For (iii), patients undergoing MSG biopsy participated in a parallel design clinical trial. Half of the patients were randomly allocated to receive treatment, which was anticipated to lead to a clinically meaningful reduction in the FS, and the remaining half was randomly allocated to receive standard care.

For the calculation of the FS, the simulations required information on both the number of foci observed within the biopsied glands and the area of the individual glands observed within each patient's biopsy. For both parameters, the choice of the simulation model and inputs was based on data from the Optimising Assessment in Sjögren's Syndrome (OASIS) cohort.

OASIS recruits new patients attending a multidisciplinary SjD clinic at Queen Elizabeth Hospital Birmingham [16–18]. Symptom duration was defined as time since onset of major sicca symptom. Routine haematoxylin and eosin diagnostic slides from 32 consecutive SjD patients meeting 2016 ACR/EULAR classification criteria were scanned with a high-resolution scanner and rescored to provide the number of foci and area for each individual MSG within the sample. Formalin-fixed paraffin-embedded (FFPE) blocks were sectioned at 3 levels approximately 30 μm apart, and the most representative section was selected for analysis. All subjects provided written informed consent, and the study was approved by the Wales Research Ethics Committee 7 formerly Dyfed Powys REC; 13/WA/0392.

Simulation models

Multilevel models were used to simulate both the number of foci and area of individual glands to incorporate variability between patients and between glands within patients. For the number of foci, multilevel Poisson and negative binomial

models were considered. A negative binomial model was preferred for simulating the log-number of foci because it fitted the OASIS data better. For the simulations on diagnosis based on FS ≥1 and reproducibility of FS, the model included:

- A regression intercept ($beta_0$), which represents the mean log-number of foci within a gland when all remaining model parameters take the value of 0.
- A patient-level random-effect parameter (u) assumed to be normally distributed around a mean of 0 with variance σ_u^2. For the reproducibility simulation, u was kept constant for both biopsies performed within patients because they were assumed to be performed successively.
- A gland-level overdispersion random-effect parameter (ω), with the exponentiated values assumed to be gamma-distributed around a mean of 1 with variance a.

For sample size calculations, an additional fixed-effect parameter was included to reflect the absolute reduction in the log-number of foci for the treatment group ($beta_1$).

The OASIS data showed a right-skewed distribution for gland area, which was best normalised using the square root function, and a strong positive relationship with the number of foci within glands. Therefore, a linear regression mixed-effects model with a square root link function was employed for simulating the area of the glands, accounting for the relationship between the area of glands and the number of foci within glands. For all 3 simulations, the model included:

- a regression intercept (β_0), which represents the mean square root of the area of glands when all remaining model parameters take the value of 0.
- a fixed-effect parameter reflecting the relationship between the number of foci and the area of the glands (β_1).
- a patient-level random-effect parameter assumed to be normally distributed around a mean of 0 with variance σ_v^2. For the reproducibility simulation, v was kept constant for both biopsies performed within patients because no systematic differences in the total area of the 2 samples taken from each patient were assumed.
- a gland-level random error term assumed to be normally distributed around a mean of 0 with variance σ_e^2.

Simulation inputs

Information on the input values used for each parameter are presented in Table 1. For the number of foci, the base-case scenario input values used for diagnosis based on FS ≥1 and reproducibility of FS were $beta_0 = 0.45$, $\sigma_u^2 = 0.64$, and $\alpha = 0.55$. Following Leckie et al [19], the expected value for the mean number of foci within glands was 2.16 (see Table 1).

From this base-case scenario, variations to the input values of $beta_0, \sigma_u^2$, and α were made individually as follows:

- Changing the input value of $beta_0$ to 0.10 and 0.80.
- Changing the input value of σ_u^2 to 0.30 and 1.32.
- Changing the input value of a to 0.30 and 1.

The same input values were used to generate the log-number of foci for both the treatment and standard care groups, while input values of −0.35, −0.51, and −0.69 were used for the mean absolute reduction of the log-number of foci in the treatment group ($beta_1$). These values correspond to a relative reduction in the number of foci within glands of 30%, 40%, and 50%, respectively. Given that all biopsies had the same number of glands within each generated data set, these relative reductions also apply to both the total number of foci observed within the biopsy and the FS.

For the area of the glands, the base-case scenario input values used for all 3 simulations were $\beta_0 = 1.35$, $\beta_1 = 0.25$, $\sigma_v^2 = 0.09$, and $\sigma_e^2 = 0.42$. The expected value for the mean square root of the area of glands was 1.89mm^2 (see Table 1).

From this base-case scenario, variations to the input values of $\beta_0, \beta_1, \sigma_v^2$, and σ_e^2 were made individually as follows:

- Changing the input value of β_0 to 1.20 and 1.50.
- Changing the input value of β_1 to 0.20 and 0.30.
- Changing the input value of σ_v^2 to 0.02 and 0.20.
- Changing the input value of $\sigma_e^2$to 0.30 and 0.56.

Table 1
Notation description, estimation method, and input values used for each simulation parameter

Description	Notation	Formula	Input values[a]
Sample size			
Number of patients	$N_{patients}$	-	15, **30**, 60
Number of glands within patients	N_{glands}	-	(2,3,4,5,6,7)[b]
Number of foci within glands			
Regression intercept[c]	$beta_0$	-	0.10, **0.45**, 0.80
Patient-level variance[c]	σ_u^2	-	0.30, **0.64**, 1.32
Gland-level variance of the exponentiated overdispersion parameter	a	-	0.30, **0.55**, 1
Mean value	$\overline{Foci}$	$e^{\left(beta_0 + \frac{\sigma_u^2}{2}\right)}$	-
Mean absolute reduction for the treatment group[c,d]	$beta_1$	-	−0.35, **−0.51**, −0.69
Mean value for the treatment group[d]	$\overline{Foci_{trt}}$	$e^{\left(beta_0+beta_1+\frac{\sigma_u^2}{2}\right)}$	-
(Square root of) area of individual glands within MSG biopsy			
Regression intercept	β_0	-	1.20, **1.35**, 1.50
Regression slope implying relationship with number of foci within glands	β_1	-	0.20, **0.25**, 0.30
Patient-level variance	σ_v^2	-	0.02, **0.09**, 0.20
Gland-level variance	σ_e^2	-	0.30, **0.42**, 0.56
Mean value	$\overline{\sqrt{Area}}$	$\beta_0 + \beta_1 \times \overline{Foci}$	-

MSG, minor salivary gland.
[a] Values in bold indicates the input value used for the base-case scenario.
[b] All 6 input values were used in all different analyses performed to examine how estimates change for different simulated gland numbers.
[c] Estimated on the logarithm scale.
[d] Only applies to the sample size calculations.

Apart from the values used for the mean reduction of the number of foci in the treatment group, all base-case inputs were equal to the estimates obtained from the analysis of the OASIS data, while the variations made on base-case inputs were based on the uncertainty (lower and upper 95% confidence bounds) observed in these estimates. For the rationale behind the number of patients, glands within each biopsy, and simulated data sets, please see Supplementary Information.

Analysis of generated data

The 2500 data sets for the diagnosis and reproducibility simulations and 3600 data sets for the sample size calculations were simulated separately for each MSG number. Each data set contained a value for (i) the number of foci observed within each gland and (ii) the square root of the area of each gland. Assessing the generation of the 2 parameters involved calculating the absolute difference between the average of the generated values, and the expected value for each parameter. The average of the generated values for the number of foci and the area of the glands was expressed as mean (SD) or median (IQR), as appropriate. The FS of each biopsy was then calculated for each patient within each generated data set, as the total number of foci observed within each biopsy over the total glandular area, multiplied by $4mm^2$. For the diagnosis simulation, the probability of FS≥1 was estimated for each generated data set as the number of patients with a FS ≥1, over the total number of patients within the dataset. The 2500 probability estimates produced for each simulated gland number were in turn summarised using the median (IQR) and were compared across the different numbers of glands assessed.

For the reproducibility of FS, the parameter was estimated for each patient as the absolute difference between the 2 successive FS and summarised within each simulated data set using the median (IQR). The 2500 estimates produced for the median absolute difference and IQR, per gland number, were in turn summarised using the mean or median, as appropriate, and compared across the 6 gland numbers assessed.

For the sample size calculations, each generated data set was analysed using a single-level negative binomial regression model, to estimate the relative reduction in mean FS for the group receiving treatment, along with the corresponding 95% confidence interval and p-value. Once all 3600 generated data sets per simulated gland number were analysed, the power was estimated as the number of times the model succeeded to detect the assumed relative reduction in FS for the treatment group (ie, p-value produced for the treatment effect parameter being <0.05), over the 3600 replications. Data simulation and analysis were initially based on 15 patients per group, and were repeated using a larger sample size until a satisfactory power estimate was obtained (80% or 90%, depending on the simulation scenario). The sample size required for estimating the relative reduction in FS with satisfactory power was then compared across the 6 simulated numbers of glands.

Finally, to evaluate how well the treatment effect parameter ($beta_1$) was estimated from the negative binomial model, we calculated bias, precision, and coverage as described in Supplementary Information.

Table 2
Baseline characteristics of OASIS cohort patients whose parameters informed the simulations

Patient-level characteristics	
Age, mean (SD)	54.40 (13.54)
Female, n (%)	30 (94%)
BMI, kg/m^2, mean (SD)	28.51 (6.09)
Symptom duration, y, median (IQR)	4.55 (2.10, 8.68)
Anti-Ro/SSA positive, n (%)	26 (81%)
Anti-La/SSB positive, n (%)	17 (53%)
IgG, g/L, mean (SD)	15.57 (4.87)
ESSDAI score, median (range)	5 (0, 17)
Unstimulated salivary flow in 5 min, median (IQR)	0.30 (0.07, 0.76)
Salivary flow ≤0.1 ml/min, n (%)	22/31 (71%)
Schirmer's ≤5mm, n (%)	21/29 (72%)
Gland-level characteristics	
Glandular area, mm^2, median (IQR)	3.28 [1.02, 5.73]
Number of glands per patient, n (%)	
- 1	1 (3%)
- 2	2 (6%)
- 3	4 (13%)
- 4	10 (31%)
- 5	3 (9%)
- 6	7 (22%)
- 7-16	5 (16%)
Number of foci per gland, n (%)	
- 0	67 (41%)
- 1	36 (21%)
- 2	17 (10%)
- 3	12 (7%)
- 4	9 (5%)
- 5	10 (6%)
- 6	6 (4%)
- 7	4 (2%)
- 8-13	4 (2%)

BMI, body mass index; ESSDAI, EULAR Sjögren's syndrome disease activity index; La/SSB, Sjögren's syndrome antigen B; OASIS, Optimising Assessment in Sjögren's Syndrome; Ro/SSA, Sjögren's syndrome antigen A.

RESULTS

Baseline characteristics that informed the simulation

Table 2 shows the baseline characteristics of the OASIS patients whose biopsies informed the simulation. The number of glands within the biopsies varied from 1 (1/32, 3%) to 16 (1/32, 3%), with the majority of biopsies containing 3 to 6 glands (24/32, 75%). The mean number of foci within the glands was 1.79 with variance 5.41, while the glands were on average 3.99 mm^2 (variance 12.89). The corresponding mean FS was 1.79.

Generation of number of foci and area of glands

For all 3 simulations, the generated estimates for the mean number of foci within glands and the mean square root of the area of glands were normally distributed for every scenario. The absolute difference between the mean of the generated estimates and the expected value was ≤0.02 for the mean number of foci and ≤0.01 for the mean square root of the area of glands across all simulated scenarios (data not shown).

Impact of gland number on achieving a FS threshold of ≥1

For the base-case scenario, the 2500 estimates produced for the probability of FS ≥1 were normally distributed for all 6 simulated numbers of glands, with the SD being lower than the prespecified value of 0.25 (see *'Number of replications'* in the Supplementary Information), allowing the parameter to be

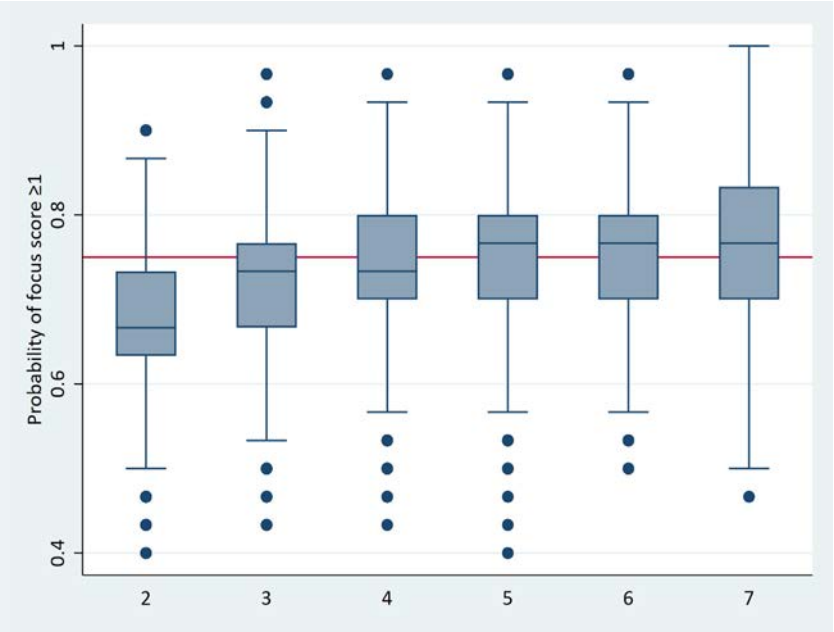


Figure 1. Probability of focus score ≥1 per simulated number of glands. A reference value of 0.75 is used to aid comparisons between different numbers of glands.

estimated with satisfactory precision. The median probability of FS ≥1 increased from 0.67 with a simulated number of 2 glands to 0.77 for a simulated number of 5 glands, with no further increase noted for a simulated number of 6 or 7 glands (Fig 1 and Supplementary Table S1). Sensitivity analyses performed with different input values remained similar to the base-case scenario, with the median probability of FS ≥1 reaching a plateau at 5 or 6 glands. The only scenario in which the median probability of FS ≥1 continued increasing up to 7 glands was when β_0 was reduced to 1.20, reflecting a lower area of the individual glands. In 2 other scenarios, no further increase was noted after 4 glands. The first of these was increasing β_0 to 1.50, which reflects a larger area of the individual glands, and the second was increasing β_1 to 0.30, which reflects a stronger positive relationship between the number of foci within glands and the area of the glands.

Impact of number of glands on reproducibility of the FS

For the base-case scenario, the 2500 estimates produced for the median absolute difference between the 2 within-patient FSs and the corresponding IQR were normally distributed for all 6 simulated numbers of glands. The SD of the median absolute difference was lower than or equal to the prespecified value of 0.25 (see *'Number of replications'* in the Supplementary Information), allowing the parameter to be estimated with satisfactory precision.

A decrease in the median absolute difference was noted for every unit increase in the number of glands; the mean of the 2500 estimates produced for the median absolute difference was 1.05 (SD = 0.25) for 2 glands, decreasing to 0.52 (SD = 0.12) for 7 glands. A similar downwards trend was observed for the IQR, with the mean of 2500 IQR estimates decreasing from 1.60 (SD = 0.48) for 2 glands to 0.71 (SD = 0.18) for 7 glands, showing that larger gland numbers allowed reproducibility to be estimated with higher precision (Fig 2 and Supplementary Table S2). Supplementary Figure shows a scatterplot of the median absolute differences in the FS against the associated IQR, illustrating that a unit increase in the number of glands leads to a reduction in both the median absolute difference and IQR values. A similar reduction was also observed in all sensitivity analyses.

Sample size calculations

For both 80% and 90% power, a significant decrease in the sample size per group required to detect a 40% relative reduction in FS for the treatment group was noted for every unit increase in the number of glands (Table 3). For 80% power, the sample size per group decreased from 62 patients with 2 glands to 25 with 7 glands. A significant increase in the required sample size was observed when the power was increased to 90%, starting from 84 patients per group with 2 glands and decreasing to 33 per group with 7 glands. Bias in the estimation of the treatment effect was present in all simulated numbers of glands for both 80% and 90% statistical power. However, this bias was less than an absolute value of 0.05, with a downwards trend for each unit increase in the number of glands. The treatment effect was estimated with satisfactory precision in all simulated scenarios (ie, SD <0.30). Undercoverage was noted for all simulated numbers of glands for both 80% and 90% statistical power. However, this was ≤3.3% across all simulated scenarios (data not shown). Reducing the assumed relative reduction for the treatment group to 30% lead to a significant increase in the sample size required, whereas a relative reduction of 50% was associated with a decrease in the required sample size (Supplementary Table S3). Both a 30% and 50% relative reduction for the treatment group yielded similar bias and coverage estimates, with a reduction in the bias and increase in the coverage noted when a larger relative reduction was assumed.

DISCUSSION

Utilising a simulation approach, we have shown that increasing the number of MSGs improves the diagnostic and measurement precision of the histological FS. For the first time, to our knowledge, we have been able to provide data to guide optimal tissue requirements. For a diagnostic threshold of FS ≥1, a minimum of 5 glands (or 4 if the glands are large) should ideally be targeted. For continuous FS values as used in clinical trials, a larger number of MSGs (eg, 6) will increase reproducibility further. Samples containing 6 or 7 glands reduced the absolute difference between 2 consecutive measurements to approximately 50% of the difference observed for samples containing only 2 glands.

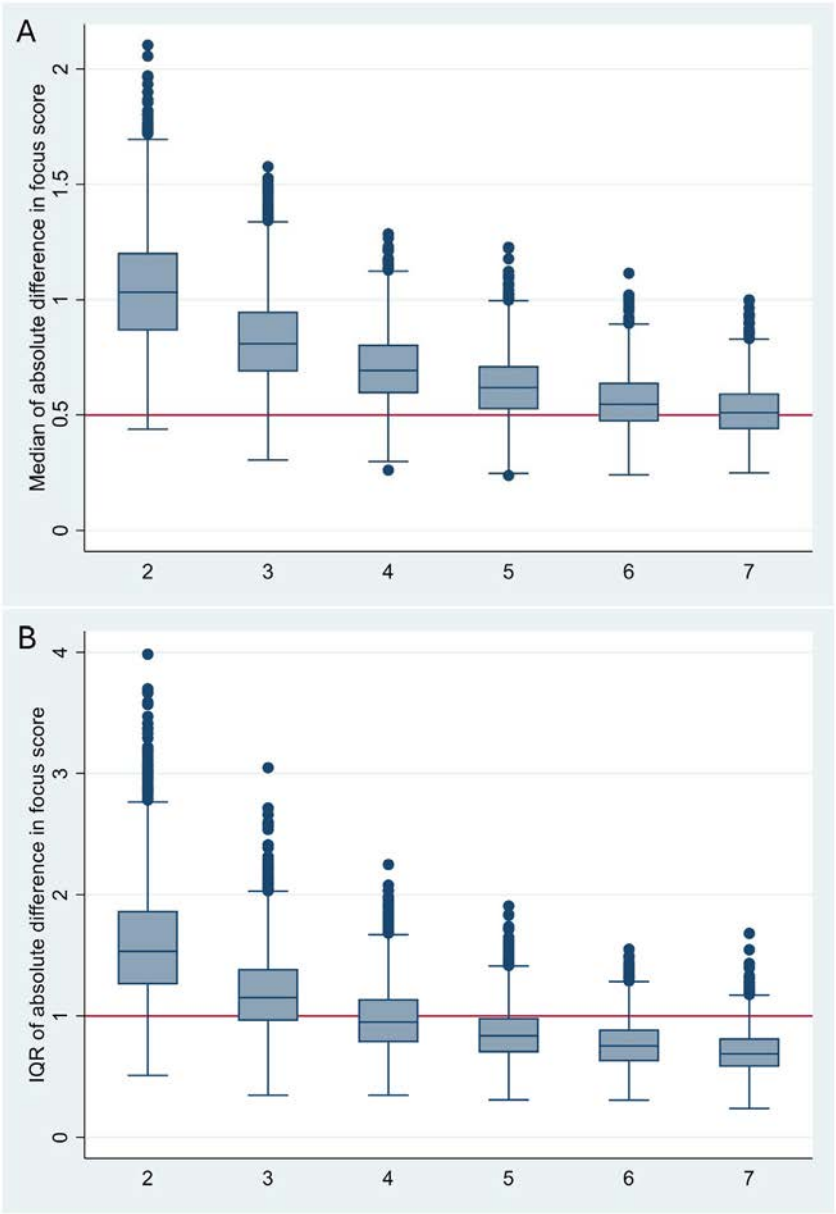


Figure 2. Impact of gland number on reproducibility of focus score. (A) Median absolute difference in focus score for each simulated number of glands. A reference value of 0.5 is used to aid comparisons between different simulated numbers of glands. (B) IQR of absolute difference in focus score for each simulated number of glands. A reference value of 1 is used to aid comparisons between different simulated numbers of glands.

Our findings have several implications. Given the importance of the FS in the diagnosis of SjD, our findings may help optimise tissue collection to ensure that samples are truly representative of the underlying glandular inflammation. This is particularly important if data continues to suggest that the FS value may be predictive of future lymphoma risk. Given that biopsy is an invasive procedure, there is ongoing interest in investigating peripheral biomarkers of salivary gland histopathology, such as CXCL13 [20,21], to aid diagnosis and monitor patient progress over time. By providing data to improve the accuracy of FS assessment, our results may assist these efforts to correlate peripheral and tissue biomarkers. Reporting summary statistics for the numbers of MSGs in biopsies may be an important addition to such research reports to help contextualise the findings. Our findings present a particular challenge for biopsy research studies requiring distribution of glands into different storage media (eg, FFPE and frozen) for different purposes, as available tissue for histology scoring is consequently reduced, which may then impact precision. Rapidly changing technologies, such as the ability to perform spatial transcriptomics at the single cell level on FFPE tissue, may help to circumvent the requirement for different storage media in the future.

We have also investigated the impact of varying MSG numbers on sample size requirements if the FS were used as a clinical outcome measure. In the absence of a large effect size, the numbers of participants required may be prohibitive, given that many trial participants may not consent to an optional biopsy substudy and a requirement for mandatory sequential biopsies may reduce overall trial recruitment. However, it is possible that mean foci size rather than a count-based measure may be more informative given that a decrease in foci size rather than a reduction in absolute number may be more achievable. Further, smaller sample sizes might still yield valuable biological insights using more data-rich approaches such as spatial transcriptomics. Our analysis considers the effect of multiple cutting levels, which has been proposed to improve the accuracy of histological assessment [22,23]. However, our analysis does not include other measures that have been suggested to improve the sensitivity or specificity of histological FS analysis for the diagnosis of SjD, such as CD45 immunostaining [24]. The analysis also does not address tissue requirements for area of infiltration, which has been proposed as a more reproducible diagnostic marker [25,26]. Composite scores that include histological parameters in addition to the FS are worthy of further

Table 3
Number of patients per group required for 80% and 90% power assuming a relative reduction of 40% for the treatment group

$N_{patients}$[a]	N_{glands}	Standard care group			Treatment group		
		Total foci[b]	Total area[b]	Focus score[c]	Total foci[b]	Total area[b]	Focus score[c]
- 80% statistical power							
62	2	4.31 (0.70)	9.55 (1.60)	1.81	2.61 (0.46)	9.59 (1.68)	1.09
42	3	6.48 (1.18)	14.35 (2.66)	1.81	3.91 (0.73)	14.37 (2.92)	1.09
35	4	8.63 (1.66)	19.11 (3.66)	1.81	5.19 (1.03)	19.12 (3.70)	1.09
30	5	10.80 (2.20)	23.81 (4.66)	1.81	6.47 (1.30)	23.85 (4.59)	1.09
28	6	13.01 (2.59)	28.79 (5.95)	1.81	7.79 (1.62)	28.82 (7.11)	1.08
25	7	15.13 (3.13)	33.46 (6.99)	1.81	9.08 (1.97)	33.49 (6.77)	1.09
- 90% statistical power							
84	2	4.30 (0.60)	9.52 (1.43)	1.81	2.59 (0.38)	9.55 (1.39)	1.09
57	3	6.49 (1.01)	14.40 (2.50)	1.80	3.90 (0.62)	14.36 (2.22)	1.09
48	4	8.65 (1.42)	19.16 (3.52)	1.81	5.20 (0.86)	19.12 (3.23)	1.09
42	5	10.83 (1.83)	23.99 (4.22)	1.81	6.46 (1.10)	23.84 (4.00)	1.08
36	6	12.87 (2.24)	28.52 (5.24)	1.81	7.75 (1.46)	28.82 (12.98)	1.08
33	7	15.10 (2.77)	33.47 (6.02)	1.81	9.08 (1.71)	33.52 (6.28)	1.08

[a] Sample size per group.
[b] Expressed as mean (SD) of mean estimates obtained from 3600 simulated data sets.
[c] Calculated as (Total foci)/(Total area) multiplied by 4.

investigation [27]. All these approaches require validation in much larger, ideally multicentre, cohort studies. With area of infiltration, typical data are heavily right-skewed, which may present challenges in developing an optimal cutoff for routine clinical use. Our data also do not address the use of parotid gland tissue, which is used in a few centres as an alternative to MSG biopsy, although published data suggest broad similarities in biological features between paired major and minor glands [28–30].

In conclusion, we have provided for the first time, to our knowledge, simulated data on the impact of MSG number on the estimate and precision of reproducibility of the histological FS. Achieving an adequate tissue sample size may facilitate both the diagnosis of SjD and the reliable quantification of an absolute FS.

Competing interests

BAF has undertaken consultancy for Novartis, BMS, Servier, Galapagos, Roche, UCB, Sanofi, Janssen, AstraZeneca, Otsuka, Amgen, and Kiniksa and received research funding from Janssen, Servier, Galapagos, and Celgene. SJB has undertaken consultancy on SjD in the past 3 years for: AbbVie, Amgen, Argenx, Aurinia, Bain, BMS, EcoR1, Iqvia, J&J/Janssen, Kiniksa, and Novartis. FB is a current employee of Candel Therapeutics. KT, SN, VP, CGS, RMB, TB, AS, MP, and JD have nothing to declare.

Acknowledgements

This work is based in part on a PhD thesis ‘Statistical Methods For Assessing The Reproducibility Of Biomarkers’ submitted to the University of Birmingham by Konstantinos Tryposkiadis (Tryposkiadis2023PhD.pdf).

Contributors

KT, FB, JD, and BAF designed the study. SN, VP, CGS, RMB, TB, SJB, FB, and BAF collected data. KT, AS, MP, and JD were responsible for statistical analysis. KT, AS, MP, JD, and BAF interpreted the data. KT and BAF drafted the manuscript. All authors critically reviewed and approved the manuscript for publication. Guarantor: BAF.

Funding

This work was funded by the National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre (NIHR 203326). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health. We also acknowledge support from the Innovative Medicines Initiative 2 Joint Undertaking (IMI 2 JU) (NECESSITY grant 806975). The Joint Undertaking received support from the European Union’s Horizon 2020 Research and Innovation Program and from the European Federation of Pharmaceutical Industries and Associations.

Patient consent for publication

All subjects provided written informed consent.

Ethics approval

Biopsy data that informed the choice of the simulation model and inputs was derived from the Optimising Assessment in Sjögren’s Syndrome study. The study was approved by the Wales Research Ethics Committee 7 formerly Dyfed Powys REC; 13/WA/0392.

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

Anonymised biopsy data that informed the simulations may be shared upon reasonable request through application to the corresponding author.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.038.

REFERENCES

[1] Nocturne G, Pontarini E, Bombardieri M, Mariette X. Lymphomas complicating primary Sjögren's syndrome: from autoimmunity to lymphoma. Rheumatology (Oxford) 2021;60(8):3513–21.

[2] Theander E, Vasaitis L, Baecklund E, Nordmark G, Warfvinge G, Liedholm R, et al. Lymphoid organisation in labial salivary gland biopsies is a possible predictor for the development of malignant lymphoma in primary Sjögren's syndrome. Ann Rheum Dis 2011;70(8):1363–8.

[3] Qin B, Wang J, Yang Z, Yang M, Ma N, Huang F, et al. Epidemiology of primary Sjögren's syndrome: a systematic review and meta-analysis. Ann Rheum Dis 2015;74(11):1983–9.

[4] Barone F, Campos J, Bowman S, Fisher BA. The value of histopathological examination of salivary gland biopsies in diagnosis, prognosis and treatment of Sjögren's syndrome. Swiss Med Wkly 2015;145:w14168.

[5] Fisher BA, Brown RM, Bowman SJ, Barone F. A review of salivary gland histopathology in primary Sjögren's syndrome with a focus on its potential as a clinical trials biomarker. Ann Rheum Dis 2015;74(9):1645–50.

[6] Fisher BA, Jonsson R, Daniels T, Bombardieri M, Brown RM, Morgan P, et al. Standardisation of labial salivary gland histopathology in clinical trials in primary Sjögren's syndrome. Ann Rheum Dis 2017;76(7):1161–8.

[7] Chatzis L, Goules AV, Pezoulas V, Baldini C, Gandolfo S, Skopouli FN, et al. A biomarker for lymphoma development in Sjogren's syndrome: salivary gland focus score. J Autoimmun 2021;121:102648.

[8] Risselada AP, Kruize AA, Goldschmeding R, Lafeber FPJG, Bijlsma JWJ, van Roon JAG. The prognostic value of routinely performed minor salivary gland assessments in primary Sjögren's syndrome. Ann Rheum Dis 2014;73 (8):1537–40.

[9] Dal Pozzolo R, Cafaro G, Perricone C, Calvacchi S, Bruno L, Colangelo A, et al. Salivary gland biopsy as a prognostic tool in Sjögren's syndrome. Expert Rev Clin Immunol 2024;20(10):1139–47.

[10] Bentley D, Fisher BA, Barone F, Kolb FA, Attley G. A randomized, double-blind, placebo-controlled, parallel group study on the effects of a cathepsin S inhibitor in primary Sjögren's syndrome. Rheumatology (Oxford) 2023;62 (11):3644–53.

[11] Haacke EA, van der Vegt B, Meiners PM, Vissink A, Spijkervet FKL, Bootsma H, et al. Abatacept treatment of patients with primary Sjögren's syndrome results in a decrease of germinal centres in salivary gland tissue. Clin Exp Rheumatol 2017;35(2):317–20.

[12] Juarez M, Diaz N, Johnston GI, Nayar S, Payne A, Helmer E, et al. A phase 2 randomized, double-blind, placebo-controlled, proof-of-concept study of oral seletalisib in primary Sjögren's syndrome. Rheumatology (Oxford) 2021;60 (3):1364–75.

[13] Mariette X, Barone F, Baldini C, Bootsma H, Clark KL, De Vita S, et al. A randomized, phase II study of sequential belimumab and rituximab in primary Sjögren's syndrome. JCI Insight 2022;7(23):e163030.

[14] Vivino FB, Gala I, Hermann GA. Change in final diagnosis on second evaluation of labial minor salivary gland biopsies. J Rheumatol 2002;29(5):938–44.

[15] Daniels TE, Cox D, Shiboski CH, Schiødt M, Wu A, Lanfranchi H, et al. Associations between salivary gland histopathologic diagnoses and phenotypic features of Sjögren's syndrome among 1,726 registry participants. Arthritis Rheum 2011;63(7):2021–30.

[16] Kollert F, Pucino V, Nayar S, Turner J, Rauz S, Richards A, et al. History of tonsillectomy is associated with glandular inflammation in Sjögren's disease. Rheumatology (Oxford) 2022;61(7) e168-70.

[17] Machowicz A, Hall I, de Pablo P, Rauz S, Richards A, Higham J, et al. Mediterranean diet and risk of Sjögren's syndrome. Clin Exp Rheumatol 2020;38 (Suppl 126(4)):216–21.

[18] Pucino V, Turner JD, Nayar S, Kollert F, Rauz S, Richards A, et al. Sjögren's and non-Sjögren's sicca share a similar symptom burden but with a distinct symptom-associated proteomic signature. RMD Open 2022;8(1):e002119.

[19] Leckie G, Browne WJ, Goldstein H, Merlo J, Austin PC. Partitioning variation in multilevel models for count data. Psychol Methods 2020;25(6):787–801.

[20] Colafrancesco S, Priori R, Smith CG, Minniti A, Iannizzotto V, Pipi E, et al. CXCL13 as biomarker for histological involvement in Sjögren's syndrome. Rheumatology (Oxford). 2020;59(1):165–70.

[21] Giacomelli R, Afeltra A, Alunno A, Bartoloni-Bocci E, Berardicurti O, Bombardieri M, et al. Guidelines for biomarkers in autoimmune rheumatic diseases - evidence based analysis. Autoimmun Rev 2019;18(1):93–106.

[22] Bredahl SK, Reibel J, Pedersen AML. Value of multilevel sectioning of labial salivary gland biopsies in the diagnosis of Sjögren's syndrome. Oral Surg Oral Med Oral Pathol Oral Radiol 2021;131(1):85–91.

[23] Morbini P, Manzo A, Caporali R, Epis O, Villa C, Tinelli C, et al. Multilevel examination of minor salivary gland biopsy for Sjogren's syndrome significantly improves diagnostic performance of AECG classification criteria. Arthritis Res Ther 2005;7(2):R343–8.

[24] Trivedi A, Cornejo KM, O'Donnell P, Dresser K, Deng A. Employing immunohistochemical staining to labial minor salivary gland biopsies from patients with Sjogren's syndrome increases diagnostic certainty. J Oral Pathol Med 2021;50(1):98–102.

[25] Donati V, Ferro F, Governato G, Fulvio G, Izzetti R, Nardini V, et al. Total area of inflammatory infiltrate and percentage of inflammatory infiltrate identify different clinical-serological subsets of primary Sjögren's syndrome better than traditional histopathological parameters. Clin Exp Rheumatol 2020;38 Suppl 126(4):195–202.

[26] Lucchesi D, Pontarini E, Donati V, Mandarano M, Sidoni A, Bartoloni E, et al. The use of digital image analysis in the histological assessment of Sjögren's syndrome salivary glands improves inter-rater agreement and facilitates multicentre data harmonisation. Clin Exp Rheumatol 2020;38 Suppl 126 (4):180–8.

[27] van Ginkel MS, Nakshbandi U, Arends S, Haacke EA, Liefers SC, Verstappen GM, et al. Increased diagnostic accuracy of the labial gland biopsy in primary Sjögren syndrome when multiple histopathological features are included. Arthritis Rheumatol 2024;76(3):421–8.

[28] Pijpe J, Kalk WW, van der Wal JE, Vissink A, Kluin PM, Roodenburg JL, et al. Parotid gland biopsy compared with labial biopsy in the diagnosis of patients with primary Sjogren's syndrome. Rheumatology (Oxford) 2007;46(2):335–41.

[29] Verstappen GM, Gao L, Pringle S, Haacke EA, van der Vegt B, Liefers SC, et al. The transcriptome of paired major and minor salivary gland tissue in patients with primary Sjögren's syndrome. Front Immunol 2021;12:681941.

[30] Nakshbandi U, van Ginkel MS, Verstappen G, Spijkervet FKL, Arends S, Haacke EA, et al. Histopathological comparison of Sjögren-related features between paired labial and parotid salivary gland biopsies of sicca patients. Rheumatology (Oxford) 2024;63(10):2670–7.

Ann Rheum Dis 84 (2025) 609–619

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Myositis

Mitochondrial transplantation as a novel therapeutic approach in idiopathic inflammatory myopathy

Jeong Yeon Kim[1], Young Cheol Kang[2], Min Jung Kim[3], Seon Uk Kim[1], Hae Rim Kang[1], Jeong Seon Yeo[2], Yujin Kim[2], Shin-Hye Yu[2], Byeongwook Song[2], Jung Wook Hwang[2], Yun-Sang Lee[4], Jung Woo Byun[4,5], Dae Hyun Yoo[6], Hyun Sook Kim[7], Kyuboem Han[2], Chun-Hyung Kim[2,*], Eun Young Lee[1,**]

[1] *Division of Rheumatology, Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea*
[2] *Paean Biotechnology Inc, Seoul, Korea*
[3] *Division of Rheumatology, Department of Internal Medicine, Seoul Metropolitan Government–Seoul National University Hospital Boramae Medical Center, Seoul, Korea*
[4] *Department of Nuclear Medicine, Seoul National University Hospital, Seoul, Korea*
[5] *Department of Biomedical Sciences, Seoul National University, Seoul, Korea*
[6] *Department of Rheumatology, College of Medicine, Hanyang University Hospital for Rheumatic Diseases, Hanyang University, Seoul, Korea*
[7] *Division of Rheumatology, Department of Internal Medicine, Soonchunhyang University Seoul Hospital, Seoul, Korea*

ARTICLE INFO

Article history:
Received 19 June 2024
Received in revised form 4 November 2024
Accepted 21 November 2024

ABSTRACT

Objectives: This study aimed to investigate the efficacy of mitochondrial transplantation as a therapeutic intervention for idiopathic inflammatory myopathy (IIM). This study used a comprehensive approach, incorporating both *in vitro* and *in vivo* IIM models, and conducted a first-in-human clinical trial to assess the effectiveness and safety of mitochondria isolated from human umbilical cord mesenchymal stem cells (PN-101).

Methods: Mitochondria isolated from umbilical cord mesenchymal stem cells were designated as PN-101. The efficacy of PN-101 was assessed using myoblasts derived from patients with IIM and C2C12 mouse perforin/granzyme B–treated myoblasts as an *in vitro* IIM model. PN-101's effect on IIM was examined using C protein–induced myositis (CIM) mice as an *in vivo* model. The efficacy and safety of PN-101 were evaluated in a phase 1/2a clinical trial involving 9 adult patients with refractory polymyositis or dermatomyositis.

Results: The myoblasts derived from patients with IIM exhibited defects in mitochondrial function and myogenesis. PN-101 transplantation enhances muscle differentiation and mitochondrial function in IIM myoblasts. PN-101 also enhanced intracellular adenosine triphosphate content, cell viability, and myogenesis in C2C12 perforin/granzyme B–treated myoblasts. In an *in vivo* model, PN-101 reduced myositis severity by exhibiting anti-inflammatory effects and restoring the CIM-induced metabolic shift. In a phase 1/2a prospective clinical trial involving adult patients with refractory IIM, PN-101 demonstrated no severe adverse drug reactions and showed at least minimal improvement in the International Myositis Assessment and Clinical Studies Group (IMACS)–Total Improvement Scores (TISs) compared with baseline.

Conclusions: PN-101 transplantation could serve as a novel treatment for IIM by enhancing mitochondrial repair and reducing inflammation in muscle tissues.

*Correspondence to Dr Chun-Hyung Kim.
**Correspondence to Eun Young Lee.
E-mail addresses: chkim@paeanbio.com (C.-H. Kim), elee@snu.ac.kr (E.Y. Lee).
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2024.11.005

WHAT IS ALREADY KNOWN ON THIS TOPIC

- High-dose glucocorticoids, either alone or in combination with immunosuppressive drugs, are the primary initial treatment for idiopathic inflammatory myopathy (IIM). However, they have serious adverse effects and limitations. Therefore, high unmet medical need for new therapy is observed.
- Patients with IIM exhibited mitochondrial damages, suggesting that mitochondria could be a potential alternative therapeutic target for IIM treatment.
- Direct application of intact functional mitochondria is becoming a promising therapeutic strategy for diseases associated with mitochondrial abnormalities.

WHAT THIS STUDY ADDS

- Defects of mitochondrial function and myogenesis were observed in *in vitro* and *in vivo* IIM models and were successfully rescued by mitochondrial transplantation.
- For the first time, an open-label, prospective 1/2a clinical research involving adult patients with refractory polymyositis and dermatomyositis showed the safety and effectiveness of mitochondrial transplantation.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

- Mitochondrial transplantation can be a therapeutic intervention as it can restore muscle strength and reduce inflammation in IIM. Thus, mitochondrial transplantation may offer a promising new treatment option for IIM.
- The use of mitochondrial transplantation therapy could be extended to treat a range of rheumatic disorders that have substantial unmet medical needs.

INTRODUCTION

Idiopathic inflammatory myopathy (IIM) is an autoimmune disorder characterised by progressive muscle weakness and loss due to immune cell-mediated myocytotoxicity [1]. The exact cause and origin of IIM remain largely unknown, limiting treatment options for patients. High-dose glucocorticoids, either alone or in combination with immunosuppressive drugs, are the primary initial treatment for IIM to suppress inflammatory responses. However, muscle weakness may persist even after inflammatory cells are cleared by conventional treatments, suggesting that nonimmune mechanisms related to muscle damage and atrophy could contribute to weakness across different disease phases [2,3]. Glucocorticoids adversely affect patients with IIM by reducing the muscle strength and mass, impairing muscle differentiation and regeneration, and causing glucocorticoid-induced atrophy [4]. Thus, developing novel therapeutic strategies to cure or slow the progression of the disease is a critical medical need.

Mitochondrial abnormalities are common in patients with IIMs, including polymyositis (PM), dermatomyositis (DM), and inclusion body myositis (IBM) [5,6]. The patients with IIM exhibited deficiencies in cytochrome C oxidase (COX) in muscle fibres [7–9], abnormalities in respiratory chain enzyme complexes [10], and mitochondrial DNA deletions [11–13]. Mitochondrial dysfunction, characterised by metabolic disturbances from a dysregulated tricarboxylic acid (TCA) cycle and excessive oxidative damage due to reactive oxygen species (ROS), has also been observed. These reports collectively suggest that mitochondria could be a potential alternative therapeutic target for IIM treatment.

Recent reports indicate that various cell types can absorb mitochondria *in vitro* [14,15] and that these can be introduced into tissues via topical or systemic injection *in vivo* [16,17]. Furthermore, systemically injected mitochondria preferentially migrate to cells and tissues with damaged mitochondria [17]. Mitotherapy, also referred to as mitochondrial transplantation therapy (MTT), involves the direct application of intact functional mitochondria and is becoming a promising therapeutic strategy for diseases associated with mitochondrial abnormalities [18,19]. In support of this claim, MTT has been successfully applied in various muscle disease models, including neuromuscular injury, acute limb ischaemia, muscular dystrophy, and sarcopenia, demonstrating that introducing healthy mitochondria into injured muscles enhances muscle mass, function, and regeneration [20–24].

In this study, we aimed to investigate the beneficial effects of MTT on IIM using mitochondria isolated fro human umbilical cord-derived mesenchymal stem cells (PN-101), mitochondria isolated from human umbilical cord–derived mesenchymal stem cells (UC-MSCs). The efficacy of PN-101 was evaluated *in vitro* using myoblasts derived from patients with IIM and *in vivo* in a C protein–induced myositis (CIM) mouse model. Furthermore, the clinical safety and efficacy of PN-101 in adults with PM or DM were initially evaluated in a phase 1/2a prospective clinical trial.

METHODS

Clinical trials

This study was conducted at Seoul National University Hospital, Hanyang University Medical Center, and Soonchunhyang University Seoul Hospital. The institutional review boards (IRBs) at each site approved the protocol. Adult patients with IIM meeting the Bohan and Peter criteria were enrolled. When the European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) criteria [25] were applied to 3 patients with PM based on the Bohan/Peter criteria, 3 patients were classified as possible PM. When the EULAR/ACR criteria were applied to 5 patients with DM based on Bohan and Peter criteria, all were classified definite DM. Patients were allowed to enter this study if they received the treatment with a prednisolone 0.5 mg/kg/d or less and/or immunosuppressive agent at stable dose for at least 4 weeks before enrollment. During the trial, these drugs were continuously maintained with a dose at enrollment. Biologic agents including intravenous immunoglobulin (IVIG) were not permitted. PN-101 was administered as a single intravenous injection at a dose of 300 mg or 600 mg to patients participated in this study. Safety, tolerability, and/or efficacy were assessed at 4, 8, and 12 weeks following PN-101 administration. This study was registered at www.ClinicalTrials.gov (NCT04976140; July 26, 2021). Muscle biopsies were obtained from consenting subjects under an IRB-approved protocol from Seoul National University Hospital (IRB number 1902-120-1013). Complete details of information including safety and efficacy evaluation, inclusion and exclusion criteria, disease duration, previous treatment, and concomitant ongoing treatment are available in Online Supplemental Tables 1 to 7.

Mitochondrial extraction

All mitochondria were produced following Good Manufacturing Practice guidelines. Human UC-MSCs cultured in Minimum

Essential Medium Eagle Alpha Modification (α-MEM; HyClone) supplemented with 10% foetal bovine serum (HyClone), 100 IU/mL penicillin, 100 mg/mL streptomycin (P/S; HyClone), and 10 ng/mL basic fibroblast growth factor (FGF-2; CHA Meditech Co). Mitochondria were isolated from UC-MSCs using differential centrifugation, with slight modifications to previously described methods [26]. Cells harvested from culture flasks were homogenised in SHE buffers (0.25 M sucrose, 20 mM HEPES, pH 7.4, 2 mM EGTA, 0.1% defatted BSA) and centrifuged at 1000 × *g* for 3 minutes at 4°C to remove cellular debris and nuclei. The supernatant was centrifuged at 12,000 × *g* for 15 minutes at 4°C to pellet the mitochondria. The pellet was resuspended in SHE buffers and centrifuged twice at 20,000 × *g* for 10 minutes each at 4°C. The resuspended pellet was then kept on ice until measurement. Mitochondrial protein concentrations were quantified using a bicinchoninic acid (BCA) assay. One milligram of PN-101 was typically obtained from 1×10^8 UC-MSCs.

RESULTS

Primary myoblasts from patients with inflammatory myositis exhibited defects in myogenesis and mitochondrial dysfunction

Initial comparisons of myotube morphology were made between primary myoblasts derived from patients with IIM and those from healthy individuals. Myotubes from IIM-derived myoblasts showed morphologic constriction by day 6 of differentiation, appearing shorter and thinner than those from normal myoblasts (Fig 1A). Immunoblot analysis measured myosin heavy chain (MHC) protein expression over a 6-day *in vitro* differentiation period. MHC protein levels were significantly lower in IIM myoblasts compared with normal myoblasts (Fig 1B). To examine the effects of inflammatory conditions on myogenesis, myoblasts were cultured for 6 days with interleukin (IL)-15, IL-6, or interferon (IFN)γ during myogenesis differentiation. IFNγ treatment, alone or combined with IL-15 or IL-6, eliminated MHC expression in normal and IIM-derived myoblasts (Fig 1C). These findings suggest that the defect in late myogenesis differentiation in myoblasts derived from patients with IIM is characterised by the absence of MHC expression.

The immunoblot assay evaluated mitochondrial protein expression in myoblasts derived from patients with IIM to detect potential mitochondrial abnormalities. A comparison of myoblasts derived from patients with IIM with normal myoblasts on day 6 of differentiation revealed decreased expression of mitochondrial oxidative phosphorylation (OXPHOS) system components, translocase of outer membrane 20 (TOM20) and Nicotinamide adenine dinucleotide reduced:ubiquinone oxidoreductase subunit B8 (NDUFB8), one of components of the mitochondrial OXPHOS system (Fig 1D-F). This indicates mitochondrial dysfunction in myoblasts derived from patients with IIM.

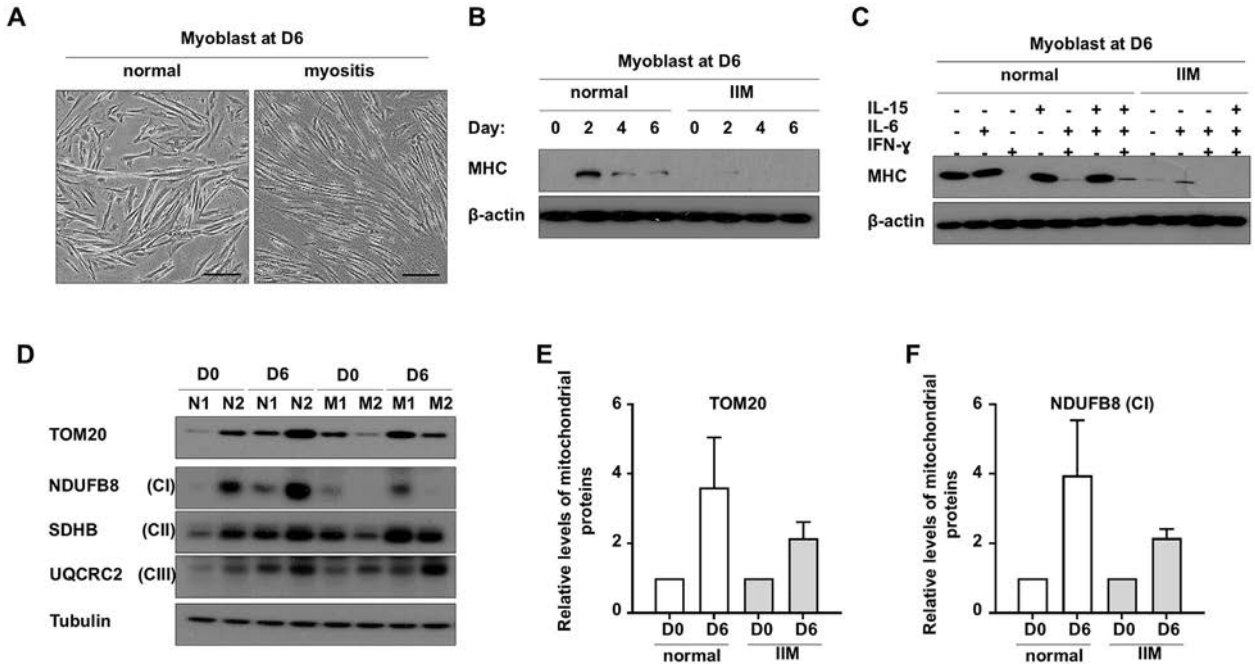


Figure 1. Defects of myogenesis and mitochondrial function in human primary myoblasts derived from patients with idiopathic inflammatory myopathy (IIM). (A) The primary myoblasts from healthy controls and patients with IIM were differentiated using 2% horse serum. Morphologic differences between control myoblasts and myoblasts derived from patients with IIM were observed at day 6 (D6) of differentiation. Scale bar = 100 mm. (B) Immunoblot analysis of myosin heavy chain (MHC) in primary myoblasts differentiated for 6 days from healthy controls and patients with IIM. Myoblasts derived from patients with IIM exhibited significantly lower MHC protein levels compared with control myoblasts. (C) Control myoblasts and myoblasts derived from patients with IIM were differentiated in 2% horse serum and treated with vehicle, interleukin-15 (IL-15), interleukin-16 (IL-6), or interferonγ (IFNγ) alone or in combination. IFNγ treatment abolished the MHC expression in both control and IIM-patient-derived myoblasts. (D) Mitochondrial protein expression in control myoblasts and myoblasts derived from patients with IIM was analysed on day 0 (D0) and D6 of differentiation using immunoblot assays. N1 and N2 are control myoblasts, while M1 and M2 are myoblasts derived from patients with IIM. UQCRC2 indicates Ubiquinol-Cytochrome c Reductase Core Protein 2. CI, CII, and CIII means the mitochondrial complex I, complex II, and complex III, respectively. (E and F) The immunoreactivity levels of TOM20 and NDUFB8 proteins were quantified and normalised to tubulin using ImageJ software (National Institutes of Health, MD, USA). The levels of TOM20 and NDUFB8 proteins in myoblasts derived from patients with IIM were reduced compared with control myoblasts. SDHB, succinate dehydrogenase iron-sulphur subunit.

Mitochondrial transplantation enhanced muscle differentiation and mitochondrial function in myositis-derived myoblasts

Given the mitochondrial dysfunction in IIM-derived myoblasts, we hypothesised that mitochondrial transplantation would restore function and enhance differentiation in myoblasts derived from patients with IIM, given their mitochondrial dysfunction. Initially, we investigated whether myoblasts could take up PN-101^{GFP} isolated from UC-MSCs infected with a lentivirus expressing mitochondrial targeting sequence (MTS)–green fluorescent protein (GFP). Coincubation of PN-101^{GFP} with primary myoblasts demonstrated successful introduction of PN-101^{GFP} into the myoblasts within 24 hours after introduction (Fig 2A).

To investigate the effects of PN-101 on muscle differentiation, myoblasts were treated with IL-15, IL-6, and IFNγ and differentiated for 6 days following PN-101 administration. PN-101 unexpectedly restored MHC expression in myoblasts, which was previously abolished by treatment with IL-15, IL-6, and IFNγ (Fig 2B). IFNγ-induced Janus kinase/signal transducers and activators of transcription (JAK-STAT) signalling activation inhibits myogenic activity [27–29]. Therefore, we investigated whether IL-15, IL-6, and IFNγ treatment induces signal transducer and activator of transcription 1 (STAT1) and phospho-STAT1 (pSTAT1) expression in myoblasts. As expected, cytokine treatment significantly enhanced pSTAT1 and STAT1 expression. However, PN-101 treatment reduced cytokine-induced pSTAT1 expression, suggesting that PN-101 may inhibit IFNγ-induced myogenic differentiation by suppressing pSTAT1 expression (Fig 2C).

Mitochondrial protein expression in myoblasts under inflammatory conditions was assessed on day 6 of differentiation to evaluate if PN-101 treatment restores mitochondrial function. IL-15, IL-6, and IFNγ treatment notably reduced NDUFB8 and TOM20 expression in myoblasts. However, PN-101 treatment can increase NUDF8B and TOM20 levels in cytokine-treated myoblasts (Fig 2D). To further explore the impact of PN-101 on mitochondrial function in myoblasts under inflammatory conditions, oxygen consumption rate (OCR) was assessed using an XF94 Flux analyser following sequential injections of oligomycin, carbonyl cyanide 4-trifluoromethoxy-phenylhydrazone (FCCP), and antimycin A. Compared with myoblasts treated solely with cytokines, PN-101 treatment enhanced basal respiration, adenine triphosphate (ATP)-linked OCR, and maximal respiration on day 6 of differentiation (Fig 2E,F). These findings collectively suggest that PN-101 transplantation lessened the

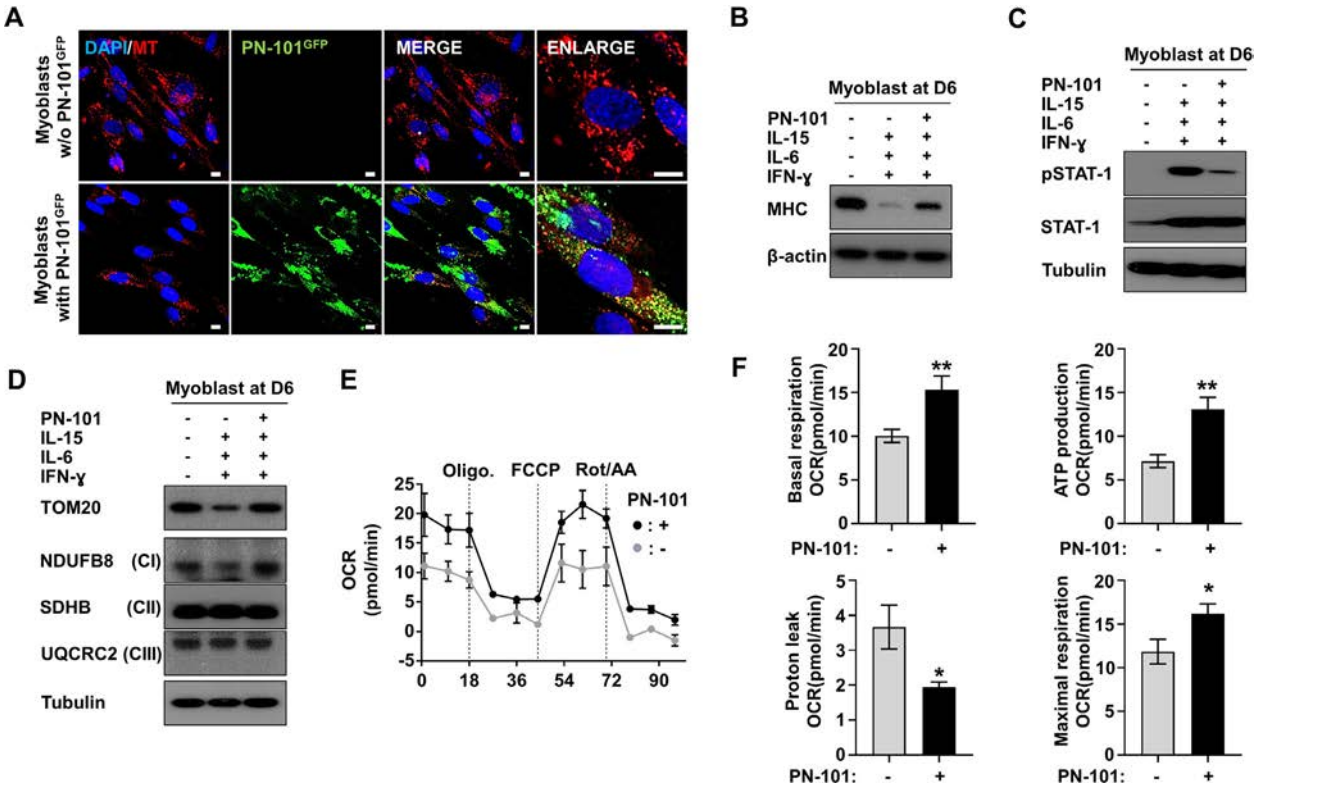


Figure 2. Effect of mitochondrial transplantation to myoblasts derived from patients with idiopathic inflammatory myopathy (IIM). (A) The transfer of fluorescence-labelled mitochondria isolated fro human umbilical cord-derived mesenchymal stem cells (PN-101) into the myoblasts. PN-101^{GFP} is derived from umbilical cord–derived mesenchymal stem cells (UC-MSCs) infected with lentivirus expressing mitochondrial targeting sequence (MTS)–green fluorescent protein (GFP). DAPI (4′,6-diamidino-2-phenylindole)-stained nuclei. The endogenous mitochondria were stained with MitoTracker Red CMXRos (ThermoFisher Scientific, MA, USA). PN-101^{GFP} was successfully introduced into myoblasts. Scale bar = 10 μm. (B) Immunoblot analysis of myosin heavy chain (MHC) in PN-101-treated primary myoblast cells from myositis patients on day 6 (D6) of differentiation. PN-101 restored MHC expression in myoblasts derived from patients with IIM treated with interferonγ (IFNγ). (C) Immunoblot analysis of signal transducer and activator of transcription 1 (STAT1) and phospho-STAT1 (pSTAT1) in myoblasts derived from patients with IIM treated with PN-101 on D6 of differentiation. PN-101 inhibits pSTAT1 expression in cytokine-treated myoblasts. (D) Immunoblot analysis of mitochondrial proteins in PN-101–treated myoblasts on D6 of differentiation. PN-101 restored TOM20 and NDUFB8 expression in cytokine-treated myoblasts. (E) Bioenergetic profiles of myoblasts treated with and without PN-101. Basal OCR values were recorded. Subsequently, 1 μM oligomycin, 1 μM carbonyl cyanide 4-trifluoromethoxy-phenylhydrazone (FCCP), and 0.5 μM rotenone/antimycin A (Rot/AA) were treated into the medium (1×10^4 cells/well). (F) Basal, Adenosine triphosphate (ATP)-linked, proton leak, and maximal oxygen consumption rates (OCR). Basal OCR was measured before oligomycin treatment, and maximum OCR was assessed following FCCP treatment, with non-mitochondrial OCR subtracted after rotenone and antimycin A treatment. OCR values were normalised to cell mass, as determined by methylene blue staining (*$P < .05$, **$P < .005$ vs without PN-101). SDHB, succinate dehydrogenase iron-sulphur subunit.

deficiency in muscle differentiation and mitochondrial damage in cytokine-stimulated myoblasts.

In the CIM mouse model, PN-101 reduced myositis severity and demonstrated anti-inflammatory effects

Intrigued by the results demonstrating improved muscle development and mitochondrial function in IIM-derived myoblasts following PN-101 transplantation, we investigated PN-101 as a potential treatment for CIM in mice, a clinically relevant murine model of PM. To assess the biodistribution of PN-101, mice received an intravenous administration of PN-101^{dsRED} isolated from UC-MSCs infected with a lentivirus expressing MTS-red fluorescent protein (dsRED). The successful introduction of PN-101 to muscle tissues was confirmed by confocal microscopy and by the detection of messenger RNA (mRNA) expression for human-specific mitochondrially encoded COX 1 (MT-CO1), as shown in Online Supplemental Figure 1A,B. To assess PN-101's efficacy, CIM mice were administered a single intravenous dose of 5 μg of PN-101 or daily dexamethasone for 7 days starting from day 7 after C protein immunisation. The mice were euthanised on day 14 (Fig 3A). The histologic analysis of the muscles by haemotoxylin and eosin (H&E) staining revealed significant mononuclear inflammatory cell infiltration in the hindlimb skeletal muscles of CIM compared with the vehicle (white arrows). Interestingly, inflammatory cell infiltrates were significantly reduced in both the PN-101 and dexamethasone treatment groups (Fig 3B). The mean histologic summation score was significantly lower in the PN-101 treatment group (0.75±0.61, n = 10) and the dexamethasone treatment group (0.5±0.63, n = 10) compared with the vehicle group (1.5±0.32, n = 7) (Fig 3C). Positron emission tomography (PET)/magnetic resonance imaging (MRI) images showed increased 18F-fluorodeoxyglucose (FDG) uptake due to CIM, which decreased following treatment with PN-101 and dexamethasone (Fig 3D). We then assessed the impact of PN-101 on inflammatory cytokines in CIM mice. In CIM mice, elevated levels of tumor necrosis factor (TNF)-α and IL-6 were observed in both serum and muscle compared with control mice. Surprisingly, PN-101 treatment significantly reduced IL-6 and TNF-α mRNA expression in CIM mouse muscle (Fig 3E) and reduced IL-6 and TNF-α protein levels in the serum of CIM mice (Fig 3F). However, treatment with dexamethasone had no effect.

Next, we assessed the muscle regeneration capacity by measuring the cross-sectional area (CSA) of myofibres in the quadriceps of CIM mice. CSA in the quadriceps of CIM mice was

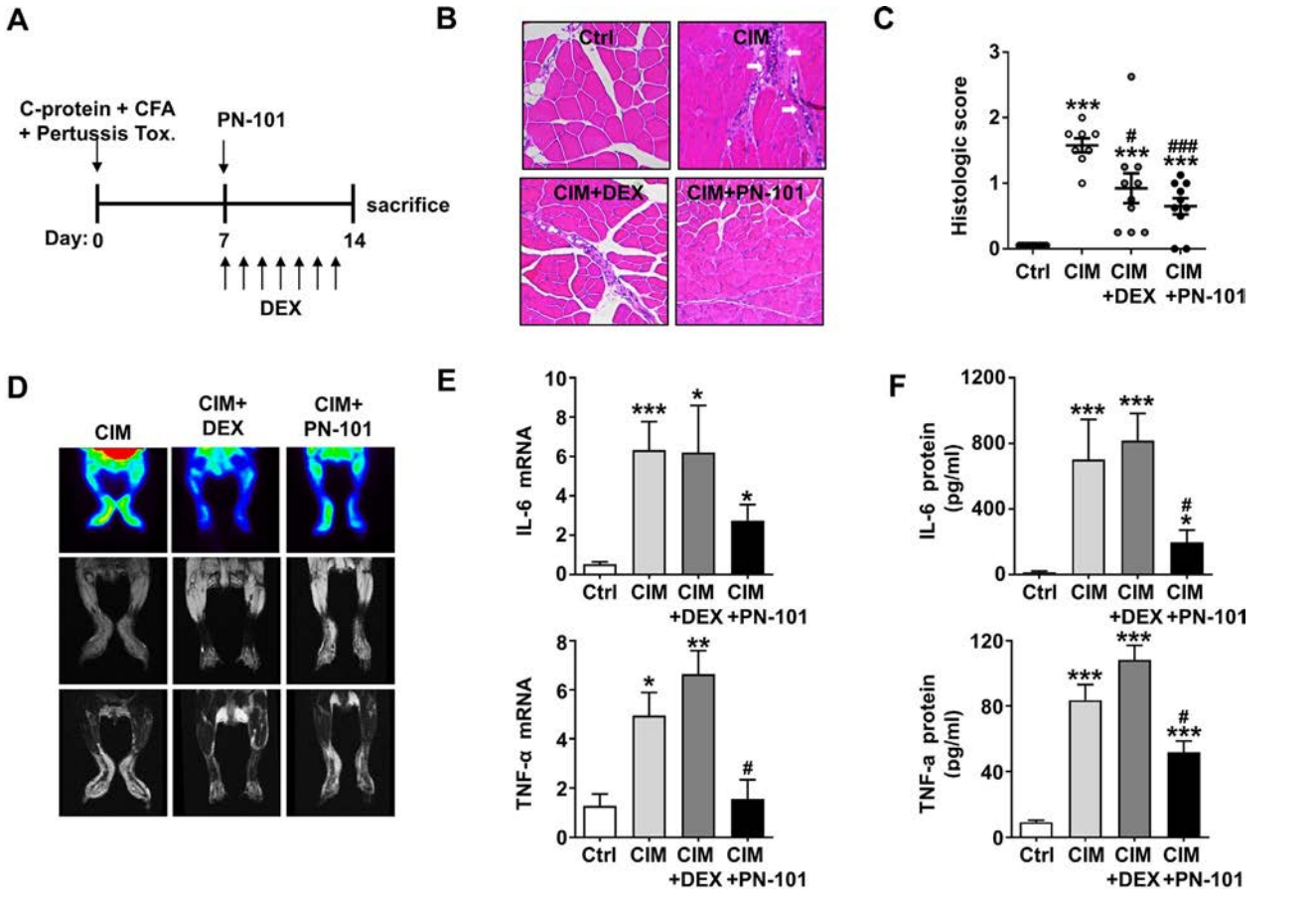


Figure 3. Anti-inflammatory effect by mitochondrial transplantation in a C protein–induced myositis (CIM) animal model. (A) Schematic diagram of animal experimental protocol for PN-101 transplantation. CIM was induced in C57BL/6 mice through immunisation with C protein fragments pertussis toxin (pertussis tox), and Complete Freund's Adjuvant (CFA). Mitochondria isolated fro human umbilical cord-derived mesenchymal stem cells (PN-101) was injected intravenously once, or dexamethasone (DEX) was administered daily for 7 days after C protein immunisation. (B) Mice were sacrificed on day 14 after immunisation, and muscle tissues were stained with haemotoxylin and eosin (H&E). Each image depicted the severity of myositis. (C) The histologic score for quadriceps was graded on a severity scale from 1 to 6 ($^{***}P < .001$ vs control [Ctrl]; $^{\#}P < .05$, $^{\#\#\#}P < .001$ vs CIM). (D) Fluorodeoxyglucose (FDG) (300 μCi) was administered intravenously, and simultaneous positron emission tomography (PET)/magnetic resonance imaging (MRI) imaging was conducted for 60 minutes using a small-animal hybrid PET/MRI scanner. (E) interferon-6 (IL-6) and tumor necrosis factor (TNF)-α messenger RNA (mRNA) levels were quantified using real time-polymerase chain reaction (RT-PCR). (F) Serum levels of IL-6 and TNF-α were measured using enzyme-linked immunosorbent assay (ELISA). Results are presented as mean ± SD ($^{*}P < .05$, $^{**}P < .005$, $^{***}P < .001$, vs Ctrl; $^{\#}P < .05$ vs CIM).

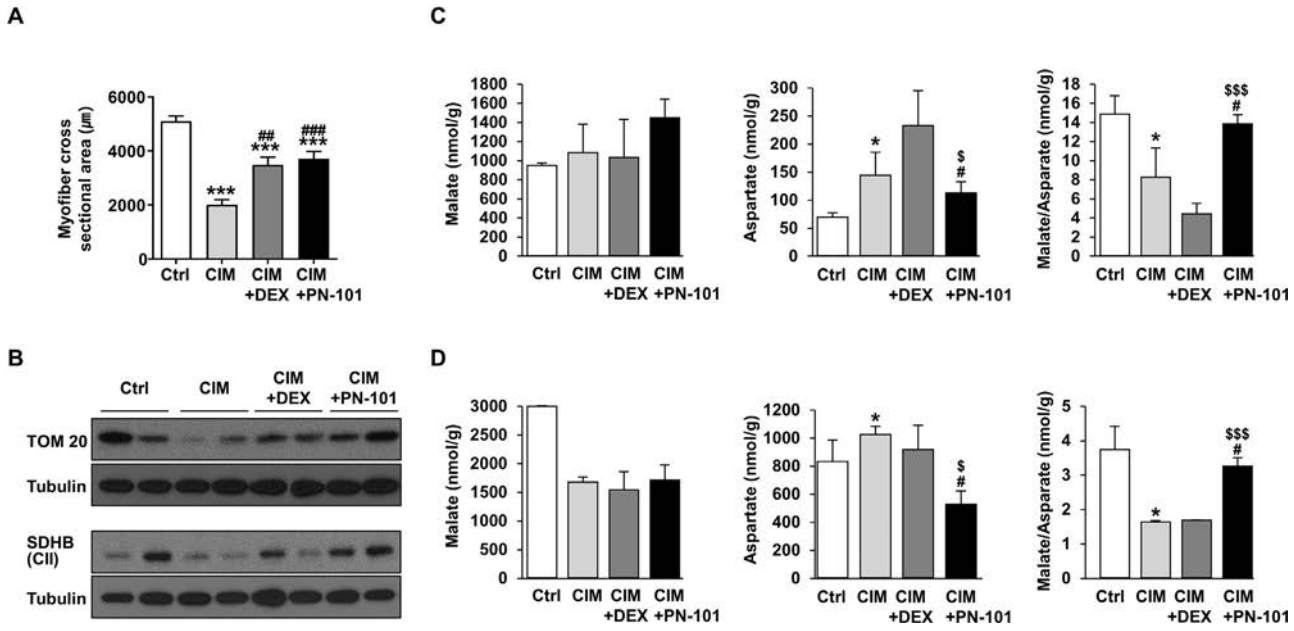


Figure 4. Mitochondrial transfer enhanced muscle mass and corrected mitochondrial defects. (A) Mitochondria isolated fro human umbilical cord-derived mesenchymal stem cells (PN-101) treatment restored normal muscle regeneration in the C protein–induced myositis (CIM) mice (***$P <$ .001, vs control [Ctrl]; ##$P <$ .005, ###$P <$ .001 vs CIM). (B) Immunoblot analysis of mitochondrial proteins in muscle tissues of the CIM mice treated with PN-101 or dexamethasone (DEX). The CIM mice showed reduced expression of mitochondrial proteins, including TOM20 and succinate dehydrogenase iron-sulphur subunit (SDHB). PN-101 treatment restored TOM20 and SDHB expression in CIM mice. (C and D) The malate/aspartate ratio was measured in the quadriceps and soleus muscles of Ctrl mice and CIM mice treated with vesicle, dexamethasone or PN-101, respectively. PN-101 treatment normalised the malate/aspartate ratio in muscle tissues of CIM mice. *P* values were calculated using a 2-tailed Student's t-test with significance markers (*$P <$ .05 vs Ctrl; #$P <$ .05 vs CIM; $$P <$ .05; $$$$P <$ .001 vs CIM + DEX).

significantly reduced compared with that of control mice, and the CIM-induced reduction of CSA was clearly restored by PN-101 treatment (Fig 4A). Mitochondrial protein expression in CIM mouse muscle tissues was also investigated. Quadriceps from CIM mice exhibited reduced expression of TOM20 and succinate dehydrogenase iron–sulphur subunit; complex II (SDHB; CII). Transplantation of PN-101 can unexpectedly restore the reduced expression of TOM20 and SDHB in CIM-induced quadriceps (Fig 4B).

Mitochondrial introduction restored the malate-aspartate shuttle defect in CIM mice

We then examined the metabolic alterations in the muscular tissues of mice with CIM. A mass spectrometry–based metabolomics study was conducted on the quadriceps and soleus muscles of control and CIM mice treated with vesicle, dexamethasone, and PN-101, respectively. Principal component analysis (PCA) was used to differentiate muscle types and treatment groups based on identified metabolites (Supplemental Fig S2A). PC1 differentiated between the quadriceps and soleus, while PC3 distinguished between control mice and those with CIM (Supplemental Fig S2A). Furthermore, they exhibited a similar pattern on the heatmap, which depicted the fold changes in metabolites (Supplemental Fig S2B). This metabolomics method enabled the identification of various compounds with differing regulatory patterns in CIM mice compared with control mice. Notably, in CIM mice, the quadriceps and soleus muscles exhibited higher aspartate levels and either lower or equivalent malate levels compared with controls. As a result, the malate/aspartate ratio significantly decreased in the quadriceps and soleus of CIM mice. PN-101 treatment significantly restored the reduced malate/aspartate ratio in CIM mice (Fig 4C,D).

PN-101 effects on mouse C2C12 myoblast cells

$CD8^+$ T cell [1] infiltration mediates muscle fibre cell death in myositis, potentially through the cytotoxic effects of T cell–secreted perforin and granzyme B [30]. We investigated the effects of PN-101 on survival, myogenesis, and mitochondrial function in C2C12 mouse myoblast cells treated with perforin and granzyme B. Initially, we confirmed the introduction of PN-101^{GFP} into C2C12 cells through direct coincubation (Fig 5A,B).

To assess the impact of perforin/granzyme B on mitochondrial function, intracellular ATP levels were measured with and without PN-101. Treatment of C2C12 cells with perforin/granzyme B, compared with controls, led to a 20% reduction in ATP levels, as shown in Figure 5C ($P <$ .001). PN-101 treatment notably restored the perforin and granzyme B–induced decrease in ATP content in a dose-dependent manner (Fig 5C).

Mitochondria are crucial in regulating programmed cell death. Therefore, we investigated the impact of PN-101 on C2C12 myoblasts subjected to perforin and granzyme B, focusing on cell death. Perforin/granzyme B induced cell death in approximately 23% of C2C12 myoblasts, as anticipated. Conversely, PN-101 treatments prevented cell death induced by perforin/granzyme B in a dose-dependent manner (Fig 5D). Furthermore, immunoblot analysis revealed elevated levels of proapoptotic proteins such as cytochrome C, caspase-3, caspase-9, and PARP were increased following the perforin/granzyme B treatment. Interestingly, the increase of proapoptotic protein expressions by perforin/granzyme B was inhibited by PN-101 treatment (Fig 5E). We then assessed the effects of PN-101 and perforin/granzyme B on myoblast differentiation. In C2C12 myoblasts treated with perforin/granzyme B, the expression of MyoD and myogenin, crucial for myoblast proliferation, decreased. However, PN-101 treatment mitigated this reduction (Fig 5F,G).

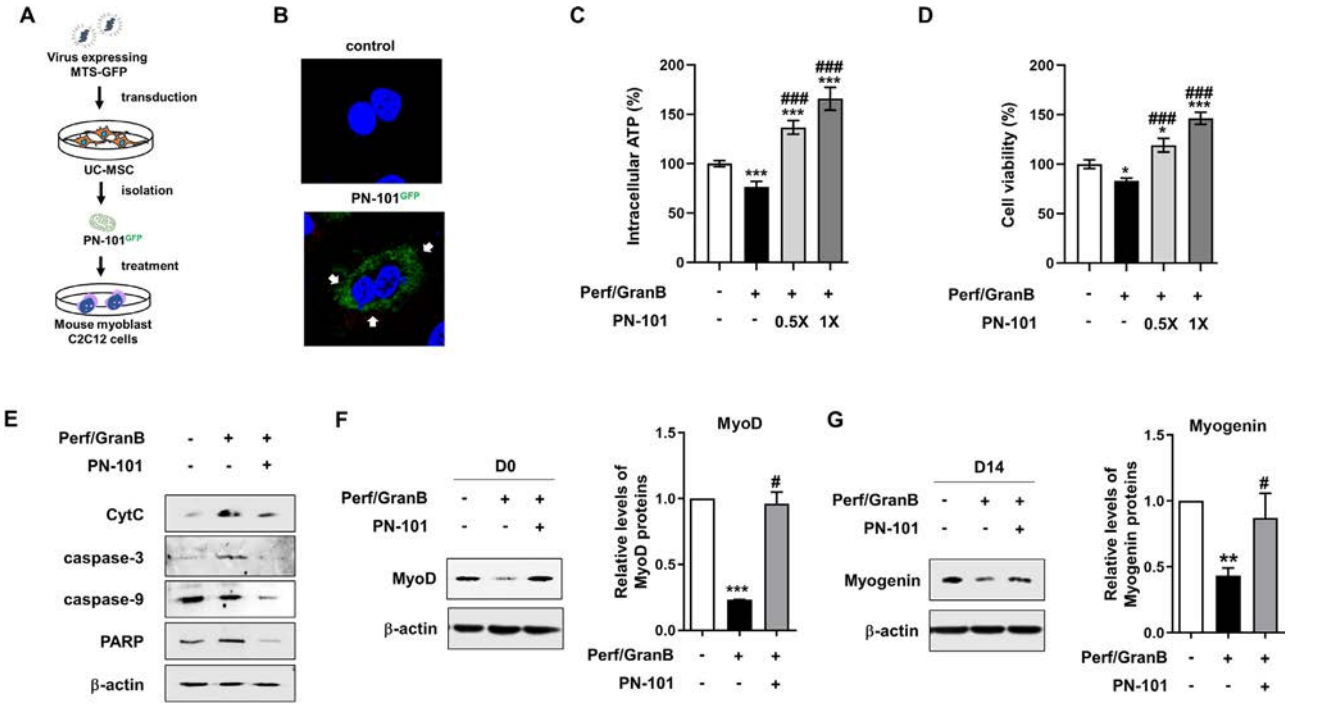


Figure 5. Effect of mitochondrial transplantation on perforin (Perf) /granzyme B (GranB)−treated C2C12 mouse myoblast cells. Perf and GranB −treated C2C12 mouse myoblast cells served as an *in vitro* model for idiopathic inflammatory myopathy (IIM). (A) Schematic diagram of the treatment of PN-101^{GFP} into the C2C12 myoblast cells. (B) Representative images of C2C12 cells treated with PN-101 ^{GFP}. White arrows indicate the successful introduction of PN-101^{GFP} into C2C12 cells. DAPI-stained nuclei. Scale bar = 20 μm. (C) The fluorescence intensity revealed the decrease of intracellular ATP levels in C2C12 cells treated with Perf/GranB. The fluorescence level in control cells was standardised to 100%. PN-101 restored and even enhanced intracellular ATP levels in C2C12 cells treated with Perf/GranB. (D) Perf/GranB treatment-induced cytotoxicity in C2C12 cells. PN-101 reduced cytotoxicity in the Perf/GranB-treated C2C12 cells. (E) Assessment of apoptosis-regulating protein expression by the immunoblot assay. Treatment with Perf/GranB elevated proapoptotic protein levels, whereas PN-101 reversed these Perf/GranB-induced alterations in apoptotic protein expression. CytC and PARP indicate cytochromeC and poly (ADP-ribose) polymerase, respectively. (F) Impact of PN-101 on MyoD expression at day 0 of differentiation. PN-101 restored the MyoD expression, which was decreased by Perf/GranB in C2C12 cells. Right panel shows the MyoD expression quantification. (G) Effect of PN-101 on myogenin expression in differentiating C2C12 cells. On day 14 of differentiation, myogenin expression in C2C12 cells decreased due to Perf/GranB treatment but was restored by PN-101. Right panel indicates the quantification of myogenin expression. Data are presented as mean ± SEM (*$P < .05$, **$P < .005$, ***$P < .001$ vs control; and $^{\#}P < .05$, $^{\#\#\#}P < .001$ vs Perf/GranB-treated group). MTS, mitochondrial targeting sequence; UC-MSC, umbilical cord−derived mesenchymal stem cell.

Patient enrollment and characteristics

Experimental results from PN-101 transplantation in both *in vitro* and *in vivo* IIM models support initiating human clinical trials. Ten patients from 3 institutions were screened following written consent. Nine subjects meeting the clinical trial criteria received the investigational product in appropriate doses for each treatment stage. The first 3 participants received 300 μg of PN-101 and successfully completed the scheduled dose limiting toxicity (DLT) evaluation. After Data and Safety Monitoring Board (DSMB) made the decision to raise the dosage, and 3 subjects were recruited to the group receiving 600 μg of PN-101 and completed the DLT evaluation on time. Three additional subjects received the 600 μg of PN-101 to further confirm its safety. However, due to injection-related adverse effects, one of the subjects received only 300 μg of PN-101. In this clinical trial, 4 participants were administered 300 μg of PN-101, and 5 subjects received 600 μg of PN-101 (Supplemental Fig S3). Subjects receiving the investigational drug were on average 49 years old (SD = 15.29) with a nearly equal sex distribution (55.56% male, 44.44% female). The study included 9 test participants: 6 with DM (66.67%) and 3 with PM (33.33%) (Supplemental Table S2).

PN-101 safety in patients with PM and DM

In the safety set, 7 of the 9 patients (77.78%) reported 19 adverse events (AEs). Fourteen AEs occurred in the 300 μg PN-101 group, compared with 5 in the 600 μg PN-101 group. Among the 19 reported AEs, treating physicians identified 4 as adverse drug reactions (ADRs) potentially related to PN-101, including increased D-dimer, headache, and injection response. The Table provides a summary of all ADRs. D-dimer levels in 2 patients (22.22%) treated with PN-101 increased but returned to baseline within 2 weeks. Headache and injection site reactions were observed as additional ADRs in 2 patients each 300 μg of PN-101, respectively (11.11%). Acute AEs such as headaches and infusion-related reactions occurred immediately after administering PN-101 but quickly resolved. Patients with AEs in both treatment groups tolerated the medication well without discontinuation. No patients experienced fatal severe AEs.

Efficacy of PN-101 in Patients with PM and DM

Efficacy was assessed by evaluating changes in the International Myositis Assessment and Clinical Studies Group (IMACS)

Table
Adverse events observed in patients of prospective investigation group during PN-101 therapy for PM/DM

MedDRA preferred term	300 μg of PN-101 (n = 4)	600 μg of PN-101 (n = 5)	Total (N = 9)
Subjects with ADRs (%) [event]	3 (75.00) [3]	19 (20.00) [1]	4 (44.44) [4]
D-dimer	1 (25.00) [1]	1 (20.00) [1]	2 (22.22) [2]
Headache	1 (25.00) [1]	0 (0.00) [0]	1 (11.11) [1]
Injection response	1 (25.00) [1]	0 (0.00) [0]	1 (11.11) [1]

ADR, adverse drug reaction; DM, dermatomyositis; PM, polymyositis.

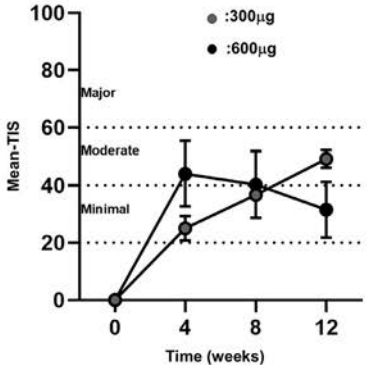
−Total Improvement Score (TIS), the IMACS-TIS response rate (≥20), and each Core Set Activity Measure of IMACS from baseline. At 4 weeks, the mean TIS for the 300μg PN-101 group was 25.0 (±6.1), and for the 600 μg PN-101 group, it was 44.0 (±22.7). Figures 6A and Supplemental Table S7 illustrate that the group treated with 600 μg of PN-101 exhibited a higher mean TIS compared with the group treated with 300 μg of PN-101 over an 8-week period. At 12 weeks, the mean (±SD) TIS was 49.2 ± 4.3 in the 300 μg of PN-101 group and was 31.5±19.6 in the 600 μg of PN-101 group, indicating that the 300 μg of PN-101 group showed continuous improvement while the 600 μg of PN-101 group exhibited reduced improvement compared with the 8th week. Nevertheless, the average TIS rose in both groups compared with the baseline. At week 12, a response of moderate improvement (TIS of ≥40) was observed in 100% of the patients (3 of 3) in the 300 μg of PN-101 group and in 20% of the patients (1 of 5) in the 600 μg of PN-101 group, and a response of minimal improvement (TIS of ≥20) and that of major improvement (TIS of ≥60) was observed in 40% (2 of 5) and 20% (1 of 5) in the 600 μg of PN-101 group, respectively. One participant in the 600 μg PN-101 group exhibited no improvement (Supplemental Fig S4). Among core set measures, it is shown that at week 12 the group treated with 300 μg of PN-101 showed about 13.4% improvement on manual muscle testing (MMT) and the group treated with 600 μg of PN-101 exhibited approximately 11.7% improvement (Fig 6B).

DISCUSSION

In this study, we demonstrated that administering PN-101 improved mitochondrial dysfunction in both myoblasts derived from patients with IIM and CIM mice. We successfully transferred exogenous fluorescence-labelled mitochondria into myoblasts *in vitro* and into muscle tissues *in vivo*. PN-101, derived from UC-MSCs, enhanced mitochondrial repair and muscle differentiation. PN-101 improved mitochondrial function, cell viability, and myogenic marker protein expression in perforin/granzyme B−treated C2C12 myoblast cells. Intravenous PN-101 administration in CIM mice decreased immune cell infiltration, suppressed proinflammatory cytokine expression, and restored CIM-induced metabolic shifts. In a 12-week prospective clinical trial, PN-101 was safe, well tolerated, and demonstrated at least minimal improvement (TIS of >20) in refractory adult patients with PM and DM compared with baseline. PN-101 transplantation may serve as a therapeutic strategy for IIM by suppressing muscle inflammation, restoring mitochondrial function, and promoting muscle regeneration.

IIM, characterised by progressive muscle weakness and inflammatory infiltrates within the muscle tissue, is an autoimmune disease [1]. The exact cause and origin of myositis remain largely unknown, limiting treatment options for patients. Treatment for myositis includes corticosteroids, glucocorticoids, and immune globulins [31]. Symptomatic treatments improve the quality of life for many patients with IIM, especially in the early stages of the disease, but they also have serious adverse effects and limitations. Recent studies have explored biologic therapies targeting TNF-α, IL-1, IL-6, IL-23, and C-X-C Motif Chemokine Ligand (CXCL)10 in IIM treatment [32−34]. The effectiveness of biological therapies, excluding rituximab, in treating IIM remains debated and their clinical use is only limited to the patients who do not effectively respond to other conventional treatments. Developing new therapeutic strategies to cure or

A



B

Core set measures	PN-101 300μg (n=3)		PN-101 600μg (n=5)	
	Baseline Mean (SD)	Week 12 Mean (SD)	Baseline Mean (SD)	Week 12 Mean (SD)
Physician global disease activity	5.1 (1.8)	2.0 (0.8)	6.2 (0.5)	5.6 (0.8)
Patient global disease activity	7.3 (1.9)	2.6 (1.8)	6.3 (1.0)	4.6 (2.6)
Extramuscular activity	2.7 (2.9)	1.0 (0.7)	4.6 (2.5)	4.3 (2.8)
MMT-8	115.3 (4.8)	130.7 (2.4)	104.4 (16.3)	116.6 (24)
HAQ	1.6 (0.6)	0.5 (0.4)	1.6 (0.5)	1.2 (0.6)
Muscle enzyme (CK)	1180.5 (184.5)	996.3 (414.3)	1506.4 (1040.2)	1215.2 (776.9)

Figure 6. Total Improvement Scores (TISs) for weeks 4, 8, and 12. (A) The mean TIS through week 12 of the open-label phase is presented. The TIS, a weighted composite score ranging from 0 to 100, reflects changes over time in 6 core measures of myositis activity. (B) Change of actual values in Core Set Measures from baseline to week 12 in the 300μg of mitochondria isolated fro human umbilical cord-derived mesenchymal stem cells (PN-101) group (n = 3) and the 600 μg of PN-101 group (n = 5), respectively. CK, creatine kinase; HAQ, Health Assessment Questionnaire; MMT, manual muscle testing.

slow the progression of the disease addresses a significant medical need.

Myogenesis directs myoblast differentiation into skeletal muscles. This process is impaired in inflammatory myopathies, including DM and IBM [35]. Our data indicated that both myoblasts derived from patients with IIM and IFNγ-treated myoblasts exhibited defective MHC expression during muscle differentiation. In the *in vitro* IIM model, perforin and granzyme B treatment led to reduced MyoD and myogenin expression in mouse myoblast C2C12 cells. These findings suggest that defective myoblast differentiation may partially contribute to the pathogenesis of myositis. This study demonstrated that PN-101 effectively restored impaired myoblast differentiation by increasing MHC expression in primary myoblasts and enhancing MyoD and myogenin expression in C2C12 cells.

Mitochondrial dysfunction is common in patients with IIM [5–7,10,12]. IIM muscle fibres exhibit mitochondrial abnormalities, including COX activity deficiency and mitochondrial DNA deletions [36]. Malate dehydrogenase and citrate synthase activities were also diminished in the muscles of patients with IIM [10]. Research demonstrated that in conditions such as ageing and disuse, unlike in IIM, elevated levels of ROS lead to oxidative damage in muscle cell components, thereby accelerating the loss of muscle mass and function [37,38]. Supporting this, the current study demonstrated mitochondrial dysfunction in both *in vitro* and *in vivo* IIM models. TOM20 and NDUFB8 expression, components of OXPHOS complex I, was significantly reduced in myoblasts derived from patients with IIM compared with normal myoblasts. Second, mouse myoblasts treated with perforin and granzyme B exhibited a significant decrease in intracellular ATP. Third, in CIM mice, muscle tissues showed reduced expression of TOM20 and SDHB, components of complex II of OXPHOS. Furthermore, metabolomic analysis revealed a significant change in the malate/aspartate ratio in the quadriceps and soleus of CIM mice. These findings suggest that mitochondrial damage in myoblasts may contribute to IIM pathophysiology. Transplanting healthy mitochondria to replace damaged ones benefits by correcting mitochondrial damage in this context [22,39–41]. Our study intriguingly revealed that PN-101 treatment can mitigate mitochondrial damage in myoblasts derived from patients with IIM and perforin/granzyme B –treated C2C12 myoblasts, as well as normalise the malate/aspartate ratio in CIM mice.

CIM mice are used as an animal model for IIM, specifically to replicate human PM [42]. PM histopathologic characteristics include primarily $CD8^+$ mononuclear T cell infiltration into nonnecrotic muscle cells in the endomysial region [43,44]. In CIM mice, $CD8^+$ T lymphocytes enriched with perforin and granzyme B effector molecules in the endomysial area, caused muscle fibre injury [42]. Various groups have demonstrated that blocking proinflammatory cytokines in the CIM mouse model could potentially serve as a therapy for treating IIM [32–34,45]. The current study demonstrated the potential of PN-101 transplantation for treating IIM in CIM mice. A single intravenous injection of PN-101 significantly mitigated CIM severity in mice by reducing inflammatory cell infiltrates, restoring diminished CSA, and attenuating proinflammatory cytokine expression in both serum and muscle tissues. PN-101 reversed mitochondrial dysfunction in CIM mice. In mammalian tissues, the malate/aspartate shuttle, crucial for regulating bioenergetics and cellular redox states, transfers reducing equivalents from the cytosol to the mitochondria for oxidation [46]. Metabolic profile alterations associated with mitochondrial abnormalities have been observed in a subset of IIM. Patients with IIM and heart failure exhibited a malate/aspartate shuttle defect [47,48]. Supporting this, metabolomic analysis revealed a deficit in the malate/aspartate shuttle in CIM mouse muscle tissues. Interestingly, PN-101 treatment significantly reduced the malate/aspartate ratio in the muscle tissues of CIM mice. Our combined *in vitro* and *in vivo* results suggest that PN-101 transplantation could be a novel therapeutic approach for IIM, enhancing myogenesis and mitochondrial function while reducing inflammatory cell infiltration and responses.

The successful outcomes of PN-101 transplantation in both *in vitro* and *in vivo* IIM models justified initiating human clinical trials. Open-label, prospective, 12-week clinical trials assessed the safety and efficacy of 2 different concentrations of PN-101 in patients with PM and DM. PN-101 was well tolerated; 9 patients experienced 19 AEs without requiring discontinuation. Additionally, PN-101 effectively improved IMACS-TISs in patients with PM and DM. A major limitation of this study is the inconclusive clinical outcomes due to the absence of a control group and a small sample size. Future studies on PN-101 for IIM treatment should be placebo-controlled, randomised, and involve a larger patient population to better assess its efficacy. Despite its limitations, this open-label, prospective clinical trial is the first proof of concept for mitochondrial transplantation in patients with PM and DM, demonstrating good safety and potential clinical efficacy as evaluated by TIS.

In conclusion, the study findings suggest that mitochondrial transplantation may effectively restore muscle strength and reduce inflammation in IIM, highlighting its potential as a therapeutic intervention. Mitochondrial transplantation may offer a promising new treatment option for IIM.

Competing interests

All authors declare they have no competing interests.

Acknowledgements

We thank all participating members of Paean Biotechnology Inc.

Contributors

J-YK and YCK: data curation, formal analysis, investigation, visualisation, methodology, writing. JSY: formal analysis, validation, methodology. YK, SUK, HRK, BS, S-HY, and JWH: data curation, methodology. MJK: data curation, writing—review and editing. HSK and DHY: formal analysis, project administration. KH: conceptualisation, resources, data curation, supervision, funding acquisition, project administration, writing—review and editing. C-HK: conceptualisation, data curation, supervision, project administration, writing—original draft, writing—review and editing. EYL: conceptualisation, resources, data curation, supervision, funding acquisition, project administration, writing—review and editing. Dr C-HK is guarantor.

Funding

This research was supported by the Korean Fund for Regenerative Medicine, funded by the Korean government (number 23D0103L1 to KH), and by Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health and Welfare, Republic of Korea (grant no. HR14C0004 to EYL). This work was also supported by the Basic Science Research Program through the National

Research Foundation of Korea (NRF) funded by the Ministry of Education NRF-2022R1A2C2006061 to CHK.

Patient consent for publication

Not applicable.

Ethics approval

This study involves human participants and was approved by the ethical committee of the Seoul National University Hospital (H-2106-091-1228), Hanyang University Hospital (HYUH-2022-02-024), and Soonchunhyang University Seoul Hospital (SCHUH-2022-05-003). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review

Not applicable.

Data availability statement

Data available upon request.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2024.11.005.

REFERENCES

[1] Dalakas MC, Hohlfeld R. Polymyositis and dermatomyositis. Lancet 2003;362(9388):971–82.
[2] Englund P, Lindroos E, Nennesmo I, Klareskog L, Lundberg IE. Skeletal muscle fibers express major histocompatibility complex class II antigens independently of inflammatory infiltrates in inflammatory myopathies. Am J Pathol 2001;159(4):1263–73.
[3] Loredo Martinez M, Zampieri S, Franco C, Ghirardello A, Doria A, Gatto M. Nonimmune mechanisms in idiopathic inflammatory myopathies. Curr Opin Rheumatol 2020;32(6):515–22.
[4] Hasselgren P-O, Alamdari N, Aversa Z, Gonnella P, Smith IJ, Tizio S. Corticosteroids and muscle wasting: role of transcription factors, nuclear cofactors, and hyperacetylation. Curr Opin Clin Nutr Metab Care 2010;13(4):423–8.
[5] Catalán-García M, García-García FJ, Moreno-Lozano PJ, Alcarraz-Vizán G, Tort-Merino A, Milisenda JC, et al. Mitochondrial dysfunction: a common hallmark underlying comorbidity between sibm and other degenerative and age-related diseases. J Clin Med 2020;9(5):1446.
[6] De Paepe B. Sporadic inclusion body myositis: an acquired mitochondrial disease with extras. Biomolecules 2019;9(1):15.
[7] Siepmann T, Tesch M, Krause F, Illigens BM, Stoltenburg-Didinger G. Polymyositis with cytochrome C oxidase negative fibers—a pathological and clinical challenge. Ann Diagn Pathol 2013;17(2):183–6.
[8] Chariot P, Ruet E, Authier FJ, Labes D, Poron F, Gherardi R. Cytochrome c oxidase deficiencies in the muscle of patients with inflammatory myopathies. Acta Neuropathol 1996;91(5):530–6.
[9] Blume G, Pestronk A, Frank B, Johns DR. Polymyositis with cytochrome oxidase negative muscle fibres. Early quadriceps weakness and poor response to immunosuppressive therapy. Brain 1997;120(Pt 1):39–45.
[10] Campos Y, Arenas J, Cabello A, Gomez-Reino JJ. Respiratory chain enzyme defects in patients with idiopathic inflammatory myopathy. Ann Rheum Dis 1995;54(6):491–3.
[11] Santorelli FM, Sciacco M, Tanji K, Shanske S, Vu TH, Golzi V, et al. Multiple mitochondrial DNA deletions in sporadic inclusion body myositis: a study of 56 patients. Ann Neurol 1996;39(6):789–95.
[12] Bhatt PS, Tzoulis C, Balafkan N, Miletic H, Tran GTT, Sanaker PS, et al. Mitochondrial DNA depletion in sporadic inclusion body myositis. Neuromuscul Disord 2019;29(3):242–6.
[13] Oldfors A, Moslemi AR, Fyhr IM, Holme E, Larsson NG, Lindberg C. Mitochondrial DNA deletions in muscle fibers in inclusion body myositis. J Neuropathol Exp Neurol 1995;54(4):581–7.
[14] Kitani T, Kami D, Kawasaki T, Nakata M, Matoba S, Gojo S. Direct human mitochondrial transfer: a novel concept based on the endosymbiotic theory. Transplant Proc 2014;46(4):1233–6.
[15] Kesner EE, Saada-Reich A, Lorberboum-Galski H. Characteristics of mitochondrial transformation into human cells. Sci Rep 2016;6:26057.
[16] Cowan DB, Yao R, Akurathi V, Snay ER, Thedsanamoorthy JK, Zurakowski D, et al. Intracoronary delivery of mitochondria to the ischemic heart for cardioprotection. PLoS One 2016;11(8):e0160889.
[17] Lee SE, Kang YC, Kim Y, Kim S, Yu SH, Park JH, et al. Preferred migration of mitochondria toward cells and tissues with mitochondrial damage. Int J Mol Sci 2022;23(24):15734.
[18] Gollihue JL, Rabchevsky AG. Prospects for therapeutic mitochondrial transplantation. Mitochondrion 2017;35:70–9.
[19] McCully JD, Cowan DB, Emani SM, Del Nido PJ. Mitochondrial transplantation: from animal models to clinical use in humans. Mitochondrion 2017;34:127–34.
[20] Turkel I, Ozerklig B, Yıldmaz M, Ulger O, Kubat GB, Tuncer M. Mitochondrial transplantation as a possible therapeutic option for sarcopenia. J Mol Med (Berl) 2023;101(6):645–69.
[21] Alway SE, Paez HG, Pitzer CR, Ferrandi PJ, Khan MM, Mohamed JS, et al. Mitochondria transplant therapy improves regeneration and restoration of injured skeletal muscle. J Cachexia Sarcopenia Muscle 2023;14(1):493–507.
[22] Orfany A, Arriola CG, Doulamis IP, Guariento A, Ramirez-Barbieri G, Moskowitzova K, et al. Mitochondrial transplantation ameliorates acute limb ischemia. J Vasc Surg 2020;71(3):1014–26.
[23] Zeng J, Liu J, Ni H, Zhang L, Wang J, Li Y, et al. Mitochondrial transplantation reduces lower limb ischemia-reperfusion injury by increasing skeletal muscle energy and adipocyte browning. Mol Ther Methods Clin Dev 2023;31:101152.
[24] Dubinin MV, Mikheeva IB, Stepanova AE, Igoshkina AD, Cherepanova AA, Semenova AA, et al. Mitochondrial transplantation therapy ameliorates muscular dystrophy in mdx mouse model. Biomolecules 2024;14(3):316.
[25] Lundberg IE, Tjärnlund A, Bottai M, Werth VP, Pilkington C, de Visser M, et al. 2017 European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Ann Rheum Dis 2017;76(12):1955–64.
[26] Yu SH, Kim S, Kim Y, Lee SE, Park JH, Cho G, et al. Human umbilical cord mesenchymal stem cell-derived mitochondria (PN-101) attenuate LPS-induced inflammatory responses by inhibiting NFκB signaling pathway. BMB Rep 2022;55(3):136–41.
[27] Shi Z, Rockey DC. Interferon-gamma-mediated inhibition of serum response factor-dependent smooth muscle-specific gene expression. J Biol Chem 2010;285(42):32415–24.
[28] Trenerry MK, Della Gatta PA, Cameron-Smith D. JAK/STAT signaling and human in vitro myogenesis. BMC Physiol 2011;11:6.
[29] Casal-Dominguez M, Pinal-Fernandez I, Mammen AL. Inhibiting interferon pathways in dermatomyositis: rationale and preliminary evidence. Curr Treatm Opt Rheumatol 2021;7(3):258–71.
[30] Cherin P, Herson S, Crevon MC, Hauw JJ, Cervera P, Galanaud P, et al. Mechanisms of lysis by activated cytotoxic cells expressing perforin and granzyme-B genes and the protein TIA-1 in muscle biopsies of myositis. J Rheumatol 1996;23(7):1135–42.
[31] Dalakas MC. Inflammatory muscle diseases. N Engl J Med. 2015;373(4):393–4.
[32] Umezawa N, Kawahata K, Mizoguchi F, Kimura N, Yoshihashi-Nakazato Y, Miyasaka N, et al. Interleukin-23 as a therapeutic target for inflammatory myopathy. Sci Rep 2018;8(1):5498.
[33] Sugihara T, Okiyama N, Watanabe N, Miyasaka N, Kohsaka H. Interleukin-1 and tumor necrosis factor alpha blockade treatment of experimental polymyositis in mice. Arthritis Rheum 2012;64(8):2655–62.
[34] Kim J, Choi JY, Park SH, Yang SH, Park JA, Shin K, et al. Therapeutic effect of anti-C-X-C motif chemokine 10 (CXCL10) antibody on C protein-induced myositis mouse. Arthritis Res Ther 2014;16(3):R126.
[35] Morosetti R, Mirabella M, Gliubizzi C, Broccolini A, De Angelis L, Tagliafico E, et al. MyoD expression restores defective myogenic differentiation of human mesoangioblasts from inclusion-body myositis muscle. Proc Natl Acad Sci U S A 2006;103(45):16995–7000.
[36] Joshi PR, Vetterke M, Hauburger A, Tacik P, Stoltenburg G, Hanisch F. Functional relevance of mitochondrial abnormalities in sporadic inclusion body myositis. J Clin Neurosci 2014;21(11):1959–63.
[37] Lightfoot AP, McArdle A, Jackson MJ, Cooper RG. In the idiopathic inflammatory myopathies (IIM), do reactive oxygen species (ROS) contribute to muscle weakness? Ann Rheum Dis 2015;74(7):1340–6.
[38] Powers SK, Hudson MB, Nelson WB, Talbert EE, Min K, Szeto HH, et al. Mitochondria-targeted antioxidants protect against mechanical ventilation-induced diaphragm weakness. Crit Care Med 2011;39(7):1749–59.

[39] Ulger O, Kubat GB. Therapeutic applications of mitochondrial transplantation. Biochimie 2022;195:1–15.
[40] McCully JD, Del Nido PJ, Emani SM. Mitochondrial transplantation for organ rescue. Mitochondrion 2022;64:27–33.
[41] Lightowlers RN, Chrzanowska-Lightowlers ZM, Russell OM. Mitochondrial transplantation-a possible therapeutic for mitochondrial dysfunction?: Mitochondrial transfer is a potential cure for many diseases but proof of efficacy and safety is still lacking. EMBO Rep 2020;21(9):e50964.
[42] Sugihara T, Sekine C, Nakae T, Kohyama K, Harigai M, Iwakura Y, et al. A new murine model to define the critical pathologic and therapeutic mediators of polymyositis. Arthritis Rheum 2007;56(4):1304–14.
[43] Arahata K, Engel AG. Monoclonal antibody analysis of mononuclear cells in myopathies. I: Quantitation of subsets according to diagnosis and sites of accumulation and demonstration and counts of muscle fibers invaded by T cells. Ann Neurol 1984;16(2):193–208.
[44] Engel AG, Arahata K. Monoclonal antibody analysis of mononuclear cells in myopathies. II: Phenotypes of autoinvasive cells in polymyositis and inclusion body myositis. Ann Neurol 1984;16(2):209–15.
[45] Okiyama N, Sugihara T, Iwakura Y, Yokozeki H, Miyasaka N, Kohsaka H. Therapeutic effects of interleukin-6 blockade in a murine model of polymyositis that does not require interleukin-17A. Arthritis Rheum 2009;60(8):2505–12.
[46] Spinelli JB, Haigis MC. The multifaceted contributions of mitochondria to cellular metabolism. Nat Cell Biol 2018;20(7):745–54.
[47] Hayes DJ, Taylor DJ, Bore PJ, Hilton-Jones D, Arnold DL, Squier MV, et al. An unusual metabolic myopathy: a malate-aspartate shuttle defect. J Neurol Sci 1987;82(1–3):27–39.
[48] Nielsen TT, Støttrup NB, Løfgren B, Bøtker HE. Metabolic fingerprint of ischaemic cardioprotection: importance of the malate-aspartate shuttle. Cardiovasc Res 2011;91(3):382–91.

Ann Rheum Dis 84 (2025) 620–631

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Systemic sclerosis

Post hoc comparison of the effectiveness of tocilizumab, rituximab, mycophenolate mofetil, and cyclophosphamide in patients with SSc-ILD from the EUSTAR database

Qingran Yan[1,2,*], Cosimo Bruni[2], Alexandru Garaiman[2], Carina Mihai[2], Suzana Jordan[2], Mike Oliver Becker[2], Muriel Elhai[2], Rucsandra Dobrota[2], Petelytska Liubov[2], Joerg Henes[3], Eric Hachulla[4], Elise Siegert[5], Alexandra Balbir-Gurman[6], Giovanna Cuomo[7], Gabriela Riemekasten[8], Stefan Heitmann[9], Davide Mohammad Reza Beigi[10], Susanne Ullman[11], Petros Sfikakis[12], Francesca Ingegnoli[13], Vera Bernardino[14], Marie-Elise Truchetet[15], Madelon Vonk[16], Francesco Del Galdo[17], Anna-Maria Hoffmann-Vold[18], Ye Shuang[1], Oliver Distler[2], EUSTAR collaborators

[1] *Department of Rheumatology, Renji Hospital, Shanghai Jiaotong University, School of Medicine, Shanghai, China*
[2] *Department of Rheumatology, University Hospital Zurich and University of Zurich, Zurich, Switzerland*
[3] *Department of Rheumatology, University Hospital Tuebingen, Tuebingen, Germany*
[4] *Department of Internal Medicine and Clinical Immunology, Referral Centre for Rare Systemic Auto-immune and Auto-inflammatory Diseases North North-West Mediterranean and Guadeloupe, University of Lille, Inserm, Unit 1286-INFINITE-Institute of Translational Research in Inflammation, Lille, France*
[5] *Department of Rheumatology, Charité-Universitätsmedizin Berlin, Berlin, Germany*
[6] *Rheumatology Institute, Rambam Health Care Campus and Rappaport Faculty of Medicine, Technion, Haifa, Israel*
[7] *Department of Precision Medicine, "Luigi Vanvitelli" University of Campania, Naples, Italy*
[8] *Department of Rheumatology, University Clinic Schleswig-Holstein (UKSH), Lubeck, Germany*
[9] *Department of Rheumatology, Marienhospital Stuttgart, Stuttgart, Germany*
[10] *Department of Clinical Internal, Anesthesiological and Cardiovascular Sciences, Sapienza University of Rome, Rome, Italy*
[11] *Department of Dermatology, University Hospital of Copenhagen, Hospital Bispebjerg, Copenhagen, Denmark*
[12] *Department of Rheumatology, National Kapodistrian University of Athens Medical School, Athens, Greece*
[13] *Department of Rheumatology, Universita' degli Studi di Milano, Milano, Italy*
[14] *Department of Internal Medicine, Autoimmune Disease Unit, Hospital Curry Cabral, Centro Hospitalar Universitário Lisboa Central, Lisbon, Portugal*
[15] *Department of Rheumatology, Bordeaux University Hospital, Bordeaux, France*
[16] *Department of Rheumatology, Radboud University Medical Center, Nijmegen, The Netherlands*
[17] *Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK*
[18] *Department of Rheumatology, Oslo University Hospital, Oslo, Norway*

*Correspondence to Dr Qingran Yan.
E-mail address: yanqingran@renji.com (Q. Yan).
Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2025.01.014

ARTICLE INFO

Article history:
Received 5 February 2024
Received in revised form 16 November 2024
Accepted 18 November 2024

ABSTRACT

Objectives: Tocilizumab (TCZ), rituximab (RTX), mycophenolate mofetil (MMF), and cyclophosphamide (CYC) are the immunosuppressants (IS) most frequently used for systemic sclerosis-associated interstitial lung disease (SSc-ILD). This post hoc study aimed to compare their effectiveness in patients with SSc-ILD from the European Scleroderma Trials and Research (EUSTAR) database.

Methods: We included radiologically confirmed SSc-ILD patients with treatment records for TCZ, RTX, MMF, or CYC. The primary endpoint was the change in forced vital capacity (FVC) percent predicted from baseline to follow-up. Analyses were adjusted for clinical and demographic characteristics, cotreatments, and follow-up duration using propensity score-based inverse probability of treatment weighting (IPTW).

Results: Nine hundred fifty-five patients with 997 treatment observations were included in the study. The median follow-up time was 11 months (IQR, 8-14 months). After IPTW, the changes in FVC percent predicted were not significantly different in the multigroup comparison ($P = .101$). Paired comparisons showed no significant difference. CYC was associated with stable FVC in logistic regression. For subgroup analysis, the treatment differences in change of FVC percent predicted among the 4 groups were not significant in patients with combination IS or previous exposure to TCZ, RTX, or conventional IS, as well as in current smokers or nonsmokers, and regardless of whether observations started either at the initiating or noninitiating stage of the treatment.

Conclusions: In this first large real-world study, the effectiveness of TCZ, RTX, MMF, and CYC on FVC change in SSc-ILD patients was not statistically different.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Mycophenolate mofetil (MMF), cyclophosphamide (CYC), tocilizumab (TCZ), and rituximab (RTX) are the most frequently used immunosuppressants (IS) for systemic sclerosis-associated interstitial lung disease (SSc-ILD). However, comparisons of their effectiveness/efficacy are limited and inconsistent.

WHAT THIS STUDY ADDS

- In this first large European Scleroderma Trials and Research (EUSTAR) real-world study, TCZ, RTX, MMF, and CYC had comparable effectiveness on SSc-ILD, although CYC showed clinically marginal but statistically significantly better effectiveness in some subanalyses.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- The findings of this study may provide a decision basis for drug selection for SSc-ILD treatment and future clinical trial design.

INTRODUCTION

Systemic sclerosis (SSc) is an autoimmune disease characterized by vasculopathy and fibrosis affecting the skin and internal organs. Among those, interstitial lung disease (ILD) is one of the most frequent complications, and strong evidence suggests that it is associated with worse outcomes [1–5]. Although remarkable advances in understanding the natural course of SSc-ILD and in the availability of new treatments have been achieved in the recent decade, the high mortality compared with other rheumatic diseases underlines that there are still large unmet clinical needs [6,7].

Mycophenolate mofetil (MMF), cyclophosphamide (CYC), tocilizumab (TCZ), and rituximab (RTX) are the immunosuppressants (IS) with the currently highest evidence for the treatment of SSc-ILD [8,9]. Among these 4, CYC was effective compared with placebo in the Scleroderma Lung Study (SLS) 1, and MMF showed comparable efficacy to CYC in SLS2, both randomized clinical trials (RCTs) designed for SSc-ILD [10,11]. TCZ was effective compared with placebo for ILD in 2 clinical trials for early, diffuse, and inflammatory SSc patients with a high risk of disease progression, although not all the subjects had ILD at baseline [12,13]. RTX was analysed in a single-country RCT in Japan and showed a significant effect compared with the placebo on forced vital capacity (FVC) [14].

However, an important limitation of these RCTs is their low representativeness of the general SSc-ILD population. For example, an analysis of the European Scleroderma Trials and Research (EUSTAR) database showed that only 1.6% of EUSTAR-registered real-life SSc-ILD patients fulfilled the modified inclusion criteria of the phase 3 TCZ trial [15]. Another issue is the lack (or inconclusive results) of head-to-head comparisons among these IS agents in SSc-ILD. Currently, the best evidence available for head-to-head comparison is the SLS2 study, showing comparable efficacy of MMF and CYC on SSc-ILD [11]. A recent single-country RCT of ILD patients with different underlying connective tissue diseases (the RECITAL trial) showed no significant difference between RTX and CYC [16]. However, the SSc subgroup was small, and definite conclusions specifically for SSc-ILD were, therefore, difficult to make [16]. On the other hand, a small single-centre RCT from India showed better efficacy of RTX compared with CYC [17].

Given these limitations and the scarce data support to guide treatment decisions in routine practice, we aimed to perform a large, real-world, observational study comparing the effectiveness of TCZ, RTX, MMF, and CYC in SSc-ILD patients. To achieve this goal, we analysed SSc-ILD patients from the EUSTAR database, with FVC change as the primary endpoint.

METHODS

Data source

EUSTAR is a large, international, multicentre, prospective registry for SSc patients [18]. For the current analysis (EUSTAR clinical project 128), data were downloaded from the EUSTAR

database in September 2022, providing information on 22,021 patients from 231 participating centres. The structure of the EUSTAR database, the minimal essential data set, and the available clinical, demographic, and diagnostic parameters have been previously described [19]. This study was performed and reported according to the Strengthening The Reporting of Observational Studies in Epidemiology (STROBE) statement [20] (Supplemental Appendix S1).

Patients and design

We included SSc patients recorded in the EUSTAR database since 2000 who fulfilled the following criteria: (1) labelled as having high-resolution computed tomography (HRCT-) or X-ray −confirmed ILD in the database; (2) having a database-recorded start date for the use of TCZ, RTX, MMF, or CYC; (3) having the end of exposure to TCZ, RTX, MMF, or CYC after the first radiological record for ILD; (4) fulfilling the 2013 classification criteria for SSc or the 1980 American College of Rheumatology classification criteria [21,22]; and (5) having at least 2 FVC records associated with the treatment period. Patients treated with any other targeted IS therapy before the start of the study agents or autologous stem cell transplantation were excluded.

The primary endpoint was the change in FVC percent predicted from baseline to the follow-up visit closest to 1 year. Secondary endpoints included absolute changes in diffusing capacity for carbon monoxide (DLCO) percent predicted, categorical changes in FVC (ΔFVC over ±3.3%, ±5%, or ±10% predicted), a composite measure for lung function progression (FVC decrease >5% predicted and DLCO decrease >15% predicted [23]), dyspnoea change assessed by the New York Heart Association (NYHA) scale, absolute changes in the modified Rodnan Skin Score (mRSS), and progression of mRSS (ΔmRSS > 5 points and ≥25% from baseline to follow-up [24]). Definitions of progression are based on the minimal clinically important difference [25,26] or consensus to define progression.

Calculation of exposure to TCZ, RTX, MMF, or CYC

We used the record of the start and end dates of TCZ, RTX, MMF, and CYC treatments from the database. For treatments with no specific information on the end date, we counted the first visit, no longer recording the treatment of interest as the end date. However, because EUSTAR is a real-life registry database recorded by investigators, we had a small proportion of patients (less than 1% of the total study population) with data unreasonable to clinical common sense (eg, 10 years of CYC treatment). These treatment periods of unreasonable duration were excluded. If multiple exposures to a given agent were recorded in 1 patient, the earliest exposure was analysed. For a patient who had been exposed to more than 1 study agent, each exposure was included separately if there was a washout interval of at least 3 months between them.

Timepoint definitions

The visit with FVC closest to the start of treatment was defined as the baseline visit, within a maximum of 4 months before the start and before the stop of treatment, as shown in Supplemental Figure S1. A minimum of 6 months follow-up was required. The follow-up visit closest to 12 months, but no later than 4 months (12 months for RTX) after the end of treatment, was chosen as the endpoint FVC.

Covariates

Covariates used in the model to calculate propensity scores included baseline and follow-up characteristics. The covariates were determined based on expert opinion. The baseline covariates included age, sex, disease duration since the first non-Raynaud's phenomenon, skin subtype of SSc, autoantibodies, baseline mRSS, current smoker, FVC percent predicted, DLCO percent predicted, ever C-reactive protein (CRP) elevation, arthritis, tendon friction rubs, joint contraction, digital ulcers (DU), history of renal crisis, myositis, left ventricular ejection fraction (LVEF < 45%), pericardial effusion, right heart catheter confirmed pulmonary hypertension (PH), reticular changes on HRCT, previous IS exposure, background antifibrotic, IS, and corticosteroid treatments and duration of follow-up.

Patient involvement

There was no specific patient involvement in this study. EUSTAR closely collaborates with patient research partners in the conduct and design of the database.

Sensitivity and subgroup analysis

The secondary analyses were not predefined. The aim of the secondary analyses was to address the clinical limitations or statistical imbalances identified in the main analysis population, serving as a supplement to strengthen the primary findings. Sensitivity analyses used (1) observations only with recorded end dates of treatment (treatments with no specific information on end date were excluded) and (2) observations since 2013, when targeted therapies were becoming more widely available. Subgroup analyses were performed as follows: IS-naive vs nonnaive patients, patients on monotherapy without combination IS vs patients on combination IS therapy, current smokers vs nonsmokers, and observations close to the start of a new therapy (initiating) vs observations with already longer treatment duration. The initiating stage was defined as the baseline visit occurring no later than the end of the sixth month after treatment initiation.

Statistical analysis

The EULAR points to consider when analysing and reporting comparative effectiveness research using observational data in rheumatology [27] were followed, with the exception of the reasons for treatment discontinuation, which were mostly unavailable in the database.

Descriptive statistics are reported as absolute numbers and relative frequencies for categorical variables and mean and SD for continuous variables. Comparison of baseline and follow-up characteristics of patients in the 4 treatment arms was performed by χ^2 test, Fisher's exact test, or the Kruskal-Wallis test, according to the distribution of the variable.

To evaluate the effectiveness of TCZ, RTX, MMF, and CYC, propensity score-based inverse probability of treatment weighting (IPTW) was used to balance the baseline and follow-up characteristics and reduce the selection bias to a minimum. The IPTW calculation used the gradient-boosted model (GBM) [28], which is designed to balance multiple groups and allow paired comparison (R package *twang* and *survey*). Paired comparisons of the study agents were performed using average treatment effects (ATEs) [28], with Bonferroni correction for *P* values. Univariate and multivariate logistic regression was used as a

supplementary method to evaluate the effectiveness of the study agents.

Missing data were imputed with multiple imputations (M = 100) using the R package *mice*. Missing information on the exposure to treatments other than study drugs, such as antifibrotic agents, was assumed as no exposure.

P values of <.05 were considered statistically significant. Analyses were performed using R 4.1 statistical software (R Development Core Team, R Foundation for Statistical Computing).

RESULTS

Cohort selection and cohort characteristics

Within the EUSTAR database, 955 patients from 103 study centres with 997 treatment observations fulfilled the inclusion criteria (Supplemental Fig S2). The missingness of each variable is described in Supplemental Table S1.

We had 4 treatment groups: the TCZ group, n = 96 (9.6%); the RTX group, n = 172 (17.3%); the MMF group, n = 511 (51.3%); and the CYC group, n = 218 (21.9%). Most of the observations (685 out of 997, 68.7%) began at the initiation stage of treatment, defined as the baseline visit occurring no later than the end of the sixth month after treatment initiation. Baseline characteristics for the TCZ, RTX, MMF, and CYC groups showed differences reflecting the indications for the different treatments (left part of Table 1, "before IPTW"). Patients from the 2 targeted therapy (TCZ and RTX) groups had higher baseline mRSS. The baseline FVC was higher in the TCZ and MMF groups. The TCZ arm had a remarkably higher proportion of arthritis and the highest dyspnoea scale at baseline. The RTX group had the highest proportion of PH and DU ever but the lowest current smoking. Patients in the RTX and CYC groups had the highest prevalence of CRP elevation ever. Disease duration was the shortest in the CYC group and the longest in the MMF group. As expected, the TCZ and RTX groups had fewer IS-naive patients and more background IS or antifibrotic treatments. Baseline age, sex, SSc subtype (limited/diffuse), and autoantibody types were not significantly different among the 4 groups. The median follow-up time for each of the 4 treatment arms ranged from 11 (CYC, TCZ, and RTX) to 12 (MMF) months (Table 1).

Effectiveness of TCZ, RTX, MMF, and CYC on lung function after IPTW

After IPTW, most of the baseline and follow-up characteristics were balanced among the 4 treatment groups, except race, combination IS use, current smoking, and honeycombing on HRCT (Table 1).

As shown in Table 2 and Figure A, after IPTW, the mean adjusted change in FVC percent predicted from baseline to follow-up was not significantly different by multiple group comparison (*P* = .101). There was also no significant difference in FVC improvement (increase over 3.3%, 5%, or 10% predicted) or worsening (decrease over 3.3%, 5%, or 10% predicted) (Table 2). In paired comparisons using ATE, the FVC change was not significantly different among the groups (Table 3). Sensitivity analyses limited to treatment observations with stopping dates recorded and to observations since 2013 (when more treatments became available) showed no statistical difference either in multigroup or paired comparisons (Supplemental Tables S2-S5). However, multivariate logistic regression showed that CYC was the only drug independently associated with maintaining or improving FVC, defined as a decline of less than 5% predicted (Supplemental Table S6).

There was no significant difference in the change of DLCO percent predicted among the 4 groups in primary or sensitivity analyses (Table 2, Fig. B, and Supplemental Tables S3 and S5). The paired comparisons were not significantly different in the primary analysis or sensitivity analyses (Table 3, Supplemental Tables S3 and S5). Progression of lung function using the composite endpoint (FVC decrease >5% predicted and DLCO decrease >15% predicted) occurred at low frequency (<5%), and we did not detect significant differences between groups (Table 2). The change in dyspnoea assessed by the NYHA scale was significantly different among the 4 treatments both before and after the IPTW (*P* < .001), mainly because the mean change in the dyspnoea scale was a little worsened in the TCZ group while slightly improved in the other 3 groups (Table 2). The 2 sensitivity analyses showed similar significant differences (Supplemental Tables S2 and S4).

Stratified analyses on FVC change

Since combination IS use was not balanced in the IPTW model for the whole study population, we stratified our analysis based on combined IS therapy (Table 4). For each study agent, no difference was observed in FVC change between patients with monotherapy and combined IS therapy. However, in cases with no combined IS therapy, the intertreatment *P* value was significant among the 4 drugs (*P* = .032), with the mean change in predicted FVC highest for the CYC arm (2.60%) and lowest for the TCZ arm (−1.38%). No paired comparisons showed significance after multiple hypothesis corrections (Table 5). The difference in FVC change between the 4 groups was not significant for patients with combination IS treatment.

Regarding stratification by previous exposure to IS, as shown in Table 4, IS-naive patients showed no significant difference in FVC change compared with non−IS-naive patients in the TCZ, RTX, and CYC groups. The effectiveness of the 4 drugs was comparable in non−IS-naive patients but was significantly different in IS-naive patients (intertreatment *P* = .002). Again, CYC showed the highest mean FVC change numerally, and TCZ showed the lowest. Further paired comparisons in IS-naive patients showed no significance (Table 5).

We also compared the FVC effectiveness of TCZ, RTX, MMF, and CYC in observations initiated at the start of treatment (within the first 6 months of treatment) vs patients already being on treatment for a longer duration. The intertreatment *P* value among the 4 agents was not significant in either initiating or noninitiating observations (Table 4). The paired comparisons were not significant in either subgroup (Table 5). Similarly, the multigroup or paired comparisons for the 4 drugs in either current smokers or noncurrent smokers were not significantly different (Tables 4 and 5).

Effectiveness of TCZ, RTX, MMF, and CYC on mRSS after IPTW

As shown in Table 2, the adjusted mean change in mRSS from baseline to follow-up was slightly decreased in all 4 groups. No statistically significant difference between treatment groups was detected in the multigroup comparison (Table 2; Fig. C) or paired comparisons (Table 3). mRSS progression occurred in less than 10% of patients of each arm, and no significant difference was detected in the multigroup comparison (Table 2). Sensitivity analyses showed no significant differences between

Table 1

Baseline and follow-up characteristics of the patients*

	Before IPTW, n (%)†					After IPTW, n (%)†				
	CYC	MMF	RTX	TCZ	Intertreatment	CYC	MMF	RTX	TCZ	Intertreatment
N	218	511	172	96	*P* values	684.97	887.31	560.3	591.68	*P* values
Demographic and patient characters										
Age (y), mean (SD)	54.78 (12.32)	55.10 (12.58)	52.33 (12.79)	54.76 (11.34)	.088	55.36 (11.75)	54.77 (12.29)	52.99 (11.66)	54.66 (11.59)	.340
Non-RP duration (y), mean (SD)	6.23 (7.29)	8.54 (7.75)	7.39 (7.55)	7.98 (9.09)	.003	7.55 (7.41)	7.97 (7.59)	7.39 (7.39)	7.88 (8.18)	.654
Female	163 (74.8)	396 (77.5)	127 (73.8)	79 (82.3)	.376	543.3 (79.3)	664.9 (74.9)	436.8 (78.0)	486.6 (82.2)	.504
dcSSc	138 (63.3)	293 (57.3)	109 (63.4)	65 (67.7)	.137	411.8 (60.1)	535.7 (60.4)	356.7 (63.7)	404.3 (68.3)	.510
Race										
Unknown	35 (16.1)	44 (8.6)	34 (19.8)	8 (8.3)	<.001	114.0 (16.6)	73.6 (8.3)	115.1 (20.5)	47.5 (8.0)	.008
White	155 (71.1)	431 (84.3)	128 (74.4)	71 (74.0)		500.4 (73.1)	746.3 (84.1)	410.8 (73.3)	458.3 (77.5)	
Black	3 (1.4)	6 (1.2)	5 (2.9)	3 (3.1)		8.9 (1.3)	13.9 (1.6)	15.5 (2.8)	14.0 (2.4)	
Asian	17 (7.8)	14 (2.7)	2 (1.2)	10 (10.4)		39.2 (5.7)	25.5 (2.9)	2.6 (0.5)	62.3 (10.5)	
Hispanic	1 (0.5)	2 (0.4)	1 (0.6)	0 (0.0)		2.7 (0.4)	2.7 (0.3)	1.9 (0.3)	0.0 (0.0)	
Others	7 (3.2)	14 (2.7)	2 (1.2)	4 (4.2)		19.8 (2.9)	25.3 (2.9)	14.4 (2.6)	9.7 (1.6)	
DU ever	83 (38.1)	184 (36.0)	92 (53.5)	40 (41.7)	.001	271.5 (39.6)	340.8 (38.4)	258.4 (46.1)	282.0 (47.7)	.353
SRC ever	3 (1.4)	11 (2.2)	6 (3.5)	4 (4.2)	.355	5.4 (0.8)	17.5 (2.0)	17.7 (3.2)	14.0 (2.4)	.345
Autoantibodies										
ACA +	27 (12.4)	85 (16.6)	16 (9.3)	12 (12.5)	.081	90.2 (13.2)	135.0 (15.2)	64.8 (11.6)	56.7 (9.6)	.567
ATA +	146 (67.0)	317 (62.0)	99 (57.6)	62 (64.6)	.275	447.5 (65.3)	555.6 (62.6)	330.3 (59.0)	425.3 (71.9)	.237
Anti-U1RNP +	10 (4.6)	37 (7.2)	13 (7.6)	3 (3.1)	.268	41.7 (6.1)	65.6 (7.4)	30.8 (5.5)	22.2 (3.8)	.645
Anti-Pol3 +	14 (6.4)	44 (8.6)	16 (9.3)	11 (11.5)	.487	35.1 (5.1)	77.9 (8.8)	40.7 (7.3)	32.6 (5.5)	.343
Anti-PM/Scl +	12 (5.5)	33 (6.5)	18 (10.5)	5 (5.2)	.196	31.9 (4.7)	58.7 (6.6)	56.2 (10.0)	14.9 (2.5)	.034
CRP elevation ever	138 (63.3)	241 (47.2)	115 (66.9)	46 (47.9)	<.001	388.7 (56.7)	457.0 (51.5)	332.4 (59.3)	280.8 (47.5)	.311
Baseline parameters										
mRSS, mean (SD)	13.09 (10.06)	10.99 (9.11)	14.07 (10.52)	14.06 (10.68)	<.001	11.93 (9.42)	12.10 (9.72)	12.48 (9.40)	12.39 (10.02)	.958
FVC% predicted, mean (SD)	78.22 (20.19)	80.59 (21.07)	76.56 (19.58)	80.56 (19.20)	.108	78.62 (19.51)	79.77 (20.60)	79.77 (20.26)	77.94 (17.67)	.812
DLCO% predicted, mean (SD)	48.45 (17.29)	52.60 (20.22)	54.18 (19.52)	55.81 (19.62)	.004	50.18 (17.96)	52.53 (19.82)	54.00 (18.36)	53.02 (17.97)	.345
Dyspnoea scale, mean (SD)	2.32 (1.05)	2.38 (1.17)	2.17 (1.09)	2.59 (1.28)	.076	2.28 (1.05)	2.41 (1.16)	2.30 (1.17)	2.69 (1.26)	.189
ESR, mm/h, mean (SD)	27.48 (22.21)	22.29 (19.91)	23.80 (21.98)	19.41 (22.68)	.005	25.37 (20.32)	22.37 (19.85)	21.23 (20.50)	19.68 (21.14)	.129
CRP, mg/L, mean (SD)	7.08 (18.07)	6.06 (17.09)	7.19 (18.05)	1.00 (1.92)	.015	8.08 (19.40)	6.01 (16.75)	6.22 (16.55)	1.01 (1.96)	.082
Arthritis	35 (16.1)	36 (7.0)	30 (17.4)	32 (33.3)	<.001	97.7 (14.3)	86.7 (9.8)	88.2 (15.7)	90.5 (15.3)	.277
Tendon friction	31 (14.2)	47 (9.2)	25 (14.5)	13 (13.5)	.106	92.4 (13.5)	87.9 (9.9)	96.5 (17.2)	55.0 (9.3)	.185
Joint contraction	67 (30.7)	164 (32.1)	82 (47.7)	44 (45.8)	<.001	213.0 (31.1)	288.7 (32.5)	223.2 (39.8)	214.8 (36.3)	.458
Myositis	28 (12.8)	34 (6.7)	23 (13.4)	9 (9.4)	.013	68.7 (10.0)	66.6 (7.5)	57.4 (10.3)	61.4 (10.4)	.73
LVEF < 45%	6 (2.8)	7 (1.4)	2 (1.2)	0 (0.0)	.266	11.8 (1.7)	12.1 (1.4)	3.0 (0.5)	0.0 (0.0)	.1
LVEF, %, mean (SD)	61.22 (8.06)	62.31 (8.62)	60.64 (9.19)	62.71 (5.91)	.062	61.02 (7.80)	62.30 (8.70)	61.38 (8.53)	62.96 (5.80)	.122
Pericardial effusion	23 (10.6)	31 (6.1)	14 (8.1)	6 (6.2)	.188	58.8 (8.6)	55.7 (6.3)	44.8 (8.0)	45.5 (7.7)	.869
RHC confirmed PH	9 (4.1)	31 (6.1)	21 (12.2)	2 (2.1)	.002	40.0 (5.8)	49.7 (5.6)	36.5 (6.5)	15.0 (2.5)	.457
Oesophageal symptoms	132 (60.6)	322 (63.0)	109 (63.4)	65 (67.7)	.686	430.9 (62.9)	555.2 (62.6)	312.4 (55.7)	419.6 (70.9)	.168
HRCT pattern										
Ground glass	134 (61.5)	307 (60.1)	94 (54.7)	55 (57.3)	.523	402.8 (58.8)	532.0 (60.0)	317.0 (56.6)	342.4 (57.9)	.939
Honeycomb	133 (61.0)	291 (56.9)	67 (39.0)	56 (58.3)	<.001	419.3 (61.2)	504.4 (56.8)	219.1 (39.1)	365.9 (61.8)	.004
Reticular change	143 (65.6)	280 (54.8)	129 (75.0)	65 (67.7)	<.001	436.7 (63.8)	532.0 (60.0)	395.1 (70.5)	360.0 (60.8)	.367
Current smoker	96 (44.0)	194 (38.0)	31 (18.0)	33 (34.4)	<.001	270.8 (39.5)	324.4 (36.6)	114.8 (20.5)	212.4 (35.9)	.017
Follow-up time, mo, median (IQR)	11 (8, 13)	12 (9, 15)	11 (7, 14)	11 (9, 13)	.013‡	11 (9, 13)	12 (8, 14)	11 (8, 13)	11 (9, 13)	.500‡

(*continued on next page*)

Table 1 (*Continued*)

	Before IPTW, n (%)†					After IPTW, n (%)†				
	CYC	**MMF**	**RTX**	**TCZ**	**Intertreatment**	**CYC**	**MMF**	**RTX**	**TCZ**	**Intertreatment**
N	218	511	172	96	***P* values**	684.97	887.31	560.3	591.68	***P* values**
Previous medications‡										
Previous IS	55 (25.2)	187 (36.6)	129 (75.0)	57 (59.4)	<.001	239.3 (34.9)	351.9 (39.7)	311.1 (55.5)	284.4 (48.1)	.017
CYC	0 (0.0)	109 (21.3)	63 (36.6)	14 (14.6)	<.001	0.0 (0.0)	207.0 (23.3)	135.0 (24.1)	95.2 (16.1)	<.001
MMF	11 (5.0)	0 (0.0)	56 (32.6)	19 (19.8)	<.001	59.9 (8.7)	0.0 (0.0)	132.7 (23.7)	102.3 (17.3)	<.001
RTX	0 (0.0)	5 (1.0)	0 (0.0)	7 (7.3)	<.001	0.0 (0.0)	9.2 (1.0)	0.0 (0.0)	35.2 (6.0)	.002
TCZ	0 (0.0)	3 (0.6)	6 (3.5)	0 (0.0)	.001	0.0 (0.0)	6.6 (0.7)	14.3 (2.6)	0.0 (0.0)	.057
AZA	23 (10.6)	66 (12.9)	17 (9.9)	5 (5.2)	.145	107.6 (15.7)	118.5 (13.4)	46.2 (8.2)	23.4 (4.0)	.018
LEF	1 (0.5)	0 (0.0)	7 (4.1)	6 (6.2)	<.001	5.9 (0.9)	0.0 (0.0)	23.4 (4.2)	18.1 (3.1)	.028
MTX	26 (11.9)	61 (11.9)	54 (31.4)	33 (34.4)	<.001	91.3 (13.3)	127.7 (14.4)	121.1 (21.6)	138.9 (23.5)	.064
No. of previous IS, mean (SD)	0.28 (0.51)	0.48 (0.71)	1.18 (0.94)	0.88 (0.91)	<.001	0.39 (0.56)	0.53 (0.75)	0.84 (0.93)	0.70 (0.88)	<.001
Background medications										
Antifibrotic agents	19 (8.7)	33 (6.5)	23 (13.4)	13 (13.5)	.014	57.1 (8.3)	61.3 (6.9)	56.0 (10.0)	65.6 (11.1)	.566
Background IS‡	8 (3.7)	11 (2.2)	34 (19.8)	25 (26.0)	<.001	29.1 (4.2)	42.3 (4.8)	64.5 (11.5)	56.4 (9.5)	.027
CYC	0 (0.0)	0 (0.0)	2 (1.2)	1 (1.0)	.042	0.0 (0.0)	0.0 (0.0)	3.0 (0.5)	1.4 (0.2)	.295
MMF	0 (0.0)	0 (0.0)	8 (4.7)	0 (0.0)	<.001	0.0 (0.0)	0.0 (0.0)	18.5 (3.3)	0.0 (0.0)	.004
RTX	0 (0.0)	2 (0.4)	0 (0.0)	0 (0.0)	.592	0.0 (0.0)	6.1 (0.7)	0.0 (0.0)	0.0 (0.0)	.495
TCZ	0 (0.0)	2 (0.4)	0 (0.0)	0 (0.0)	.592	0.0 (0.0)	6.1 (0.7)	0.0 (0.0)	0.0 (0.0)	.495
AZA	4 (1.8)	5 (1.0)	8 (4.7)	4 (4.2)	.014	16.9 (2.5)	18.0 (2.0)	11.9 (2.1)	7.3 (1.2)	.829
MTX	4 (1.8)	4 (0.8)	17 (9.9)	15 (15.6)	<.001	12.2 (1.8)	18.2 (2.0)	32.6 (5.8)	34.4 (5.8)	.067
Any steroids	153 (70.2)	334 (65.4)	131 (76.2)	54 (56.2)	.004	449.8 (65.7)	589.9 (66.5)	377.6 (67.4)	341.8 (57.8)	.449
Prednisone > 10 mg	9 (4.1)	13 (2.5)	6 (3.5)	1 (1.0)	.427	26.4 (3.8)	28.8 (3.2)	13.9 (2.5)	1.5 (0.3)	.068

* Clinical manifestations are defined according to EUSTAR standards unless otherwise indicated [19].

† Unless otherwise specified.

‡ IS included conventional IS and 2 targeted agents, tocilizumab and rituximab; other targeted agents were excluded from this study.

ACA, anticentromere antibody; Anti-Pol3, anti-RNA polymerase III antibody; ATA, antitopoisomerase I antibody; AZA, azathioprine; CRP, C-reactive protein; CYC, cyclophosphamide; dcSSc, diffused cutaneous systemic sclerosis; DLCO, diffusing capacity of the lungs for carbon monoxide; DU, digital ulcer; ESR, erythrocyte sediment rate; FVC, forced vital capacity; HRCT, high-resolution computed tomography; IPTW, propensity score-based inverse probability of treatment weighting; IS, immunosuppressant; LEF, leflunomide; LVEF, left ventricular ejection fraction; MMF, mycophenolate mofetil; mRSS, modified Rodnan Skin Score; MTX, methotrexate; PH, pulmonary hypertension; RHC, right heart catheter; RP, Raynaud's phenomenon; RTX, rituximab; SRC, scleroderma renal crisis; TCZ, tocilizumab.

Table 2
Multigroup comparison of tocilizumab, rituximab, mycophenolate mofetil, and cyclophosphamide effectiveness on primary and secondary endpoints

	Before IPTW					After IPTW				
	CYC	**MMF**	**RTX**	**TCZ**	**Intertreatment**	**CYC**	**MMF**	**RTX**	**TCZ**	**Intertreatment**
n	218	511	172	96	***P* values**	684.97	887.31	560.3	591.68	***P* values**
Change in FVC% predicted (mean, SD)	2.60 (11.21)	0.77 (12.05)	0.08 (10.93)	−1.34 (11.35)	.028*	2.44 (10.50)	0.98 (11.93)	0.49 (10.15)	−1.51 (12.46)	.101*
FVC categorical change, n (%)										
Increase >3.3% predicted	90 (41.3)	163 (31.9)	52 (30.2)	32 (33.3)	.041†	272.4 (39.8)	286.8 (32.3)	196.8 (35.1)	200.3 (33.9)	.429†
Between −3.3% to +3.3% predicted	83 (38.1)	182 (35.6)	67 (39.0)	33 (34.4)		274.1 (40.0)	321.6 (36.2)	211.2 (37.7)	200.2 (33.8)	
Decrease >3.3% predicted	45 (20.6)	166 (32.5)	53 (30.8)	31 (32.3)		138.5 (20.2)	278.9 (31.4)	152.3 (27.2)	191.2 (32.3)	
Increase >5% predicted	75 (34.4)	127 (24.9)	40 (23.3)	21 (21.9)	.010†	225.3 (32.9)	223.0 (25.1)	149.9 (26.8)	142.3 (24.1)	.099†
Between −5% to +5% predicted	111 (50.9)	255 (49.9)	91 (52.9)	47 (49.0)		367.4 (53.6)	444.0 (50.0)	285.8 (51.0)	266.7 (45.1)	
Decrease >5% predicted	32 (14.7)	129 (25.2)	41 (23.8)	28 (29.2)		92.3 (13.5)	220.3 (24.8)	124.6 (22.2)	182.7 (30.9)	
Increase >10% predicted	39 (17.9)	70 (13.7)	18 (10.5)	10 (10.4)	.034†	115.4 (16.8)	125.2 (14.1)	62.1 11.1)	81.7 (13.8)	.411†
Between −10% to +10% predicted	166 (76.1)	372 (72.8)	131 (76.2)	76 (79.2)		532.3 (77.7)	648.2 (73.1)	442.7 (79.0)	444.6 (75.1)	
Decrease >10% predicted	13 (6.0)	69 (13.5)	23 (13.4)	10 (10.4)		37.3 (5.4)	113.9 (12.8)	55.5 (9.9)	65.5 (11.1)	
Change in DLCO% predicted, mean (SD)	9.82 (25.50)	8.78 (26.02)	1.98 (21.76)	6.72 (20.88)	.008*	10.18 (26.01)	8.55 (25.41)	5.11 (20.98)	5.33 (19.64)	.227*
Composite lung function worsening,‡ n (%)	8 (3.7)	12 (2.3)	9 (5.2)	4 (4.2)	.287†	25.4 (3.7)	19.2 (2.2)	26.1 (4.6)	25.5 (4.3)	.560†
Change in dyspnoea scale,§ mean (SD)	−0.11 (0.91)	−0.19 (0.72)	−0.18 (1.01)	0.56 (1.20)	<.001¶	−0.13 (0.91)	−0.22 (0.76)	−0.29 (1.14)	0.57 (1.22)	<.001¶
Change in mRSS, mean (SD)	−1.67 (6.76)	−1.80 (6.90)	−2.07 (7.48)	−3.93 (8.97)	.052*	−1.20 (6.68)	−2.07 (7.23)	−1.03 (7.01)	−3.09 (8.23)	.125*
mRSS worsening,‖ n (%)	15 (6.9)	30 (5.9)	14 (8.1)	7 (7.3)	.753†	53.6 (7.8)	50.8 (5.7)	55.5 (9.9)	31.3 (5.3)	.422†

* Analysis of variance test.
† $\chi 2$ test or Fisher's exact test.
‡ FVC decrease >5% predicted and DLCO decrease >15% predicted.
§ New York Heart Association (NYHA) scale.
‖ Increase of mRSS > 5 points and ≥25% from baseline to follow-up.
¶ Kruskal-Wallis test.

CYC, cyclophosphamide; DLCO, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; IPTW, propensity score-based inverse probability of treatment weighting; MMF, mycophenolate mofetil; mRSS, modified Rodnan Skin Score; RTX, rituximab; TCZ, tocilizumab.

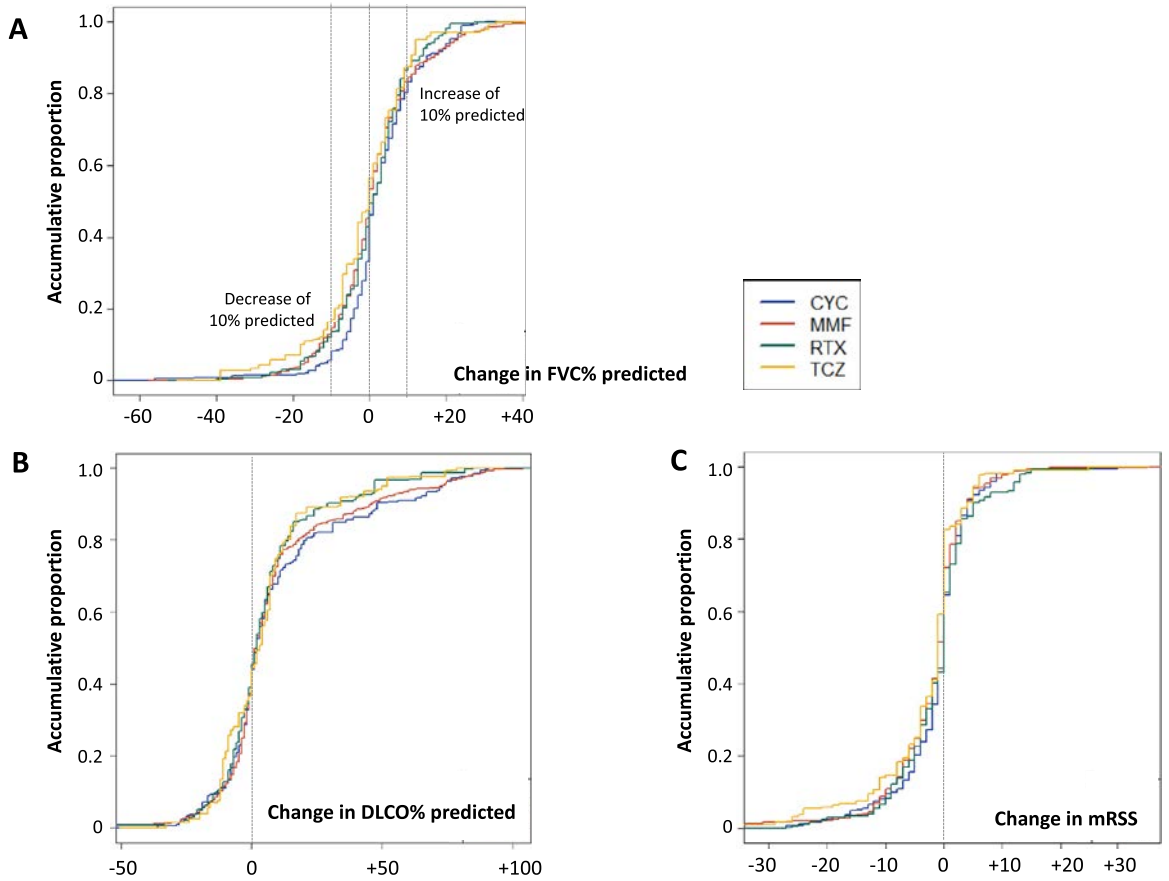


Figure. The accumulative proportion of changes in (A) forced vital capacity (FVC%) predicted, (B) diffusing capacity of the lungs for carbon monoxide (DLCO) predicted, and (C) modified Rodnan Skin Score (mRSS; after propensity score-based inverse probability of treatment weighting). CYC, cyclophosphamide; DLCO, diffusing capacity of the lungs for carbon monoxide; MMF, mycophenolate mofetil; RTX, rituximab; TCZ, tocilizumab.

treatment groups in mRSS outcomes either (Supplemental Tables S2-S5).

DISCUSSION

This is the first real-world study comparing the effectiveness of TCZ, RTX, MMF, and CYC in SSc-ILD. We addressed both initiating and long-standing treatments, with the majority being in the initiating phase. We first described the prescription pattern of these 4 IS agents for SSc-ILD in clinical practice. As expected, MMF and CYC were mostly used as the first-line treatment, and the targeted therapies were used as second-line treatments. TCZ was used more frequently in SSc patients with arthritis, implying arthritis was a major indication for TCZ selection by physicians in practice. Unexpectedly, the TCZ group did have rather infrequent CRP elevation, indicating that inflammation was not the major reason for TCZ selection in clinical practice. It must be noted that most of the TCZ patients in the EUSTAR registry were treated before the results of the phase 3 TCZ trial were published, which included early inflammatory patients and where arthritis was not an inclusion criterion. Moreover, in some countries, TCZ was only available for patients with polyarthritis.

We did not find differences in the multigroup comparison between the effectiveness of TCZ, RTX, MMF, and CYC on FVC change in patients with SSc-ILD. This was the primary endpoint of our study and is consistent with results from the 2 available large RCTs, SLS2 and RECITAL, which did not find differences when comparing the efficacy of CYC with MMF and RTX, respectively. Other RCTs comparing the remaining therapies are unavailable, and our effectiveness study provides the first data on their comparisons.

Interestingly, several subanalyses of our study point to a potentially higher effectiveness of CYC vs the other agents, mainly TCZ. These included subgroup analyses and logistic regression, especially for patients with monotherapy and IS-naive patients. CYC and TCZ have never been compared in SSc-ILD, and this potentially better effectiveness of CYC was observed for the first time. There are several potential explanations for this phenomenon. First, although TCZ has shown protective efficacy on FVC in the phase 3 trial for early inflammatory diffused cutaneous SSc patients, these convincing effects have not been tested in SSc-ILD patients with other non-inflammatory phenotypes. Our analysis comes from a real-life cohort, which is not as enriched and selected as in RCTs, and effects might be different in different subgroups of patients with SSc-ILD. Second, there are also no standard dose regimens like in RCTs, as they are selected by the treating physicians. We would like to emphasize that this result of potentially better effectiveness of CYC vs TCZ should be interpreted with great caution. All differences were derived from secondary analysis; the differences in FVC among the 4 agents were small and usually below the minimal clinically important difference [26] for SSc-ILD patients; the NYHA dyspnoea scale poorly discriminates between NYHA stage 2 and 3 in clinical routine practice. Therefore, our findings need confirmation in other cohorts and prospective RCTs on representative cohorts.

Combination IS treatment, especially a combination of biologics and conventional IS, is common practice in SSc-ILD, and background treatment with conventional IS as the -standard of

Table 3
Paired comparisons of tocilizumab, rituximab, mycophenolate mofetil, and cyclophosphamide effectiveness*

Change in FVC% predicted from baseline to follow-up

	Estimate mean difference	SE	95% CI			Adjusted *P* value†
MMF vs CYC	−1.46	0.92	−3.27	to	0.36	.462
CYC vs TCZ	3.94	1.81	0.40	to	7.48	.116
MMF vs TCZ	2.49	1.73	−0.91	to	5.89	.606
CYC vs RTX	1.95	1.16	−0.32	to	4.23	.372
MMF vs RTX	0.49	1.05	−1.56	to	2.55	1.000
TCZ vs RTX	−2.00	1.87	−5.66	to	1.67	1.000

Change in DLCO% predicted from baseline to follow-up

	Estimate mean difference	SE	95% CI			Adjusted *P* value
MMF vs CYC	−1.63	2.32	−6.17	to	2.91	1.000
CYC vs TCZ	4.86	3.07	−1.17	to	10.88	.458
MMF vs TCZ	3.23	2.59	−1.86	to	8.31	.856
CYC vs RTX	5.07	3.03	−0.87	to	11.00	.378
MMF vs RTX	3.44	2.54	−1.54	to	8.42	.705
TCZ vs RTX	0.21	3.24	−6.15	to	6.57	1.000

Change in mRSS from baseline to follow-up

	Estimate mean difference	SE	95% CI			Adjusted *P* value
MMF vs CYC	−0.87	0.61	−2.07	to	0.32	.602
CYC vs TCZ	1.90	1.01	−0.08	to	3.87	.241
MMF vs TCZ	1.02	0.95	−0.84	to	2.88	1.000
CYC vs RTX	−0.16	0.79	−1.71	to	1.38	1.000
MMF vs RTX	−1.04	0.71	−2.44	to	0.36	.583
TCZ vs RTX	−2.06	1.07	−4.17	to	0.05	.222

* Calculated by average treatment effects of the propensity score-based inverse probability of treatment weighting model.
† With Bonferroni's correction.CYC, cyclophosphamide; DLCO, diffusing capacity of the lungs for carbon monoxide; FVC, forced vital capacity; MMF, mycophenolate mofetil; mRSS, modified Rodnan Skin Score; RTX, rituximab; TCZ, tocilizumab.

care is allowed in a growing number of SSc RCTs [12,14,29]. However, background IS might mask the potential efficacy of the study agent, which is a concern in the design of RCTs. An example is the SENSCIS (Safety and Efficacy of Nintedanib in Systemic Sclerosis) trial for SSc-ILD [30]. In the SENSCIS trial, the treatment difference in the nintedanib vs placebo arm was more prominent in background MMF− patients than in MMF+ patients, as measured by achieving an FVC increase ≥3% predicted at week 52 (the odds ratio of treatment vs placebo was 1.01 [95% CI, 0.57-1.81] in the MMF+ group and 3.17 [95% CI, 1.63-6.16] in the MMF− group; subgroup heterogeneity $P = .011$) [31]. Here, in this study with a large sample size, we also showed that the treatment difference among TCZ, RTX, MMF, and CYC was significant in monotherapy patients but not in patients with combination IS therapy. Moreover, comparisons of a specific agent with or without IS cotreatment showed no significant additional benefits with background IS in our study. These data emphasize again the challenges of background treatments for the design of RCTs in SSc.

This study had some limitations. Safety analyses were not available because these data were not captured in the EUSTAR database. For example, mortality status in the database could not be verified independently. We found that 113/955 patients had recorded death. However, most of the deaths occurred very far from the observation period. In addition, the completeness of follow-up was unclear, and patients without follow-up could, in part, be patients who died. Only 27 deaths were relevant as they happened during treatment or within 6 months after stopping the drugs of interest. Among them, 11 (2.2%) deaths were associated with MMF, 5 (2.3%) with CYC, 10 (5.8%) with RTX, and 1 (1.0%) with TCZ. Safety analysis is relevant because safety is a major reason for the selection between CYC and MMF in clinical practice [9], with CYC having a somewhat higher number of adverse events. We mostly used FVC to assess treatment response in this study but could not include additional outcomes to reflect ILD progression, such as the patient's experience and radiological changes, as recommended by current guidelines [32]. As an inherited shortage of registry data, many patients were excluded because of a lack of follow-up or unmatched measurements with treatments. Although this exclusion may cause bias, our study population had features similar to those of the whole EUSTAR ILD population [15] and other SSc-ILD cohorts [33]. The present study could overestimate the true exposure period of TCZ, RTX, MMF, and CYC. Forty-six percent of the observations did not have information for the treatment end date; thus, they had to be estimated. However, the start date of treatment was available in all observations, the end date was available in most cases, and the missed end date was conservatively estimated. Patients dying after baseline FVC measurement were excluded as they did not fulfil the inclusion criteria for the second FVC measurement. Similarly, patients with a severe disease course might have missed their follow-up FVC because of their deteriorating condition. This might have been different between the treatment groups and could have led to different results in their effectiveness. In addition, this study included a population of mostly Caucasian European patients. Different responses to treatment in different ethnicities, drug availability, and insurance policies in different countries and periods may

Table 4
Multigroup comparison of subgroup changes in forced vital capacity percent predicted from baseline

	CYC n = 684.97*	MMF n = 887.31	RTX n = 560.3	TCZ n = 591.68	Intertreatment *P* value
Background IS					
Background IS−	2.60 (10.50)†	1.00 (12.08)	0.43 (10.32)	−1.38 (12.94)	.032
Background IS+	−1.10 (10.37)	0.73 (8.59)	0.93 (8.89)	−2.74 (6.62)	.402
Intersubgroup *P* value	.16	.92	.84	.26	
Previous IS exposure‡					
IS-naive patients	2.60 (11.35)	0.30 (11.94)	0.62 (9.16)	−1.38 (10.73)	.002
Previous IS+	2.14 (8.77)	2.02 (11.86)	0.38 (10.95)	−1.64 (14.20)	.927
Intersubgroup *P* value	.25	.07	.93	.35	
Initiating treatment					
Baseline in the first 6 mo of treatment	2.46 (10.83)	1.00 (11.40)	0.51 (10.33)	−2.05 (13.08)	.065
Baseline after the first 6 mo of treatment	2.39 (5.13)	0.96 (13.06)	0.32 (1.44)	−0.57 (9.41)	.848
Intersubgroup *P* value	.38	.69	.73	.66	
Current smoking					
Current smoker	2.60 (9.76)	1.12 (12.12)	−1.89 (9.75)	0.33 (13.71)	.160
Noncurrent smoker	2.34 (11.59)	0.90 (11.61)	1.10 (11.44)	−2.54 (9.74)	.270
Intersubgroup *P* value	.83	.91	.33	.91	

* After propensity score-based inverse probability of treatment weighting.
† Mean (SD).
‡ Previous IS exposure included cyclophosphamide, mycophenolate mofetil, rituximab, tocilizumab, azathioprine, leflunomide, and methotrexate.
CYC, cyclophosphamide; IS, immunosuppressant; MMF, mycophenolate mofetil; RTX, rituximab; TCZ, tocilizumab.

Table 5
Paired comparisons of subgroup change in forced vital capacity percent predicted from baseline*

	Estimated mean difference	SE	95% CI			Adjusted *P* value	Estimated mean difference	SE	95% CI			Adjusted *P* value
	Background IS− patients						**background IS+ patients**					
MMF vs CYC	−1.60	0.95	−3.46	to	0.26	.368	1.83	3.77	−5.55	to	9.21	1.000
CYC vs TCZ	3.97	1.98	0.10	to	7.85	.178	1.64	3.24	−4.70	to	7.98	1.000
MMF vs TCZ	2.37	1.91	−1.37	to	6.11	.856	3.47	2.72	−1.86	to	8.81	.825
CYC vs RTX	2.17	1.24	−0.27	to	4.60	.328	−2.03	3.42	−8.74	to	4.68	1.000
MMF vs RTX	0.56	1.13	−1.66	to	2.78	1.000	−0.20	2.94	−5.96	to	5.56	1.000
TCZ vs RTX	−1.81	2.07	−5.87	to	2.25	1.000	−3.67	2.22	−8.03	to	0.68	.411
	IS-naive patients						**Previous IS+ patients**					
MMF vs CYC	−2.29	1.16	−4.57	to	−0.02	.194	−0.13	1.53	−3.13	to	2.88	1.000
CYC vs TCZ	3.98	2.03	0.01	to	7.95	.200	3.79	3.10	−2.29	to	9.86	.890
MMF vs TCZ	1.69	1.93	−2.10	to	5.47	1.000	3.66	2.97	−2.17	to	9.48	.877
CYC vs RTX	1.97	1.83	−1.62	to	5.56	1.000	1.77	1.60	−1.36	to	4.89	1.000
MMF vs RTX	−0.32	1.73	−3.70	to	3.06	1.000	1.64	1.33	−0.98	to	4.25	.882
TCZ vs RTX	−2.01	2.40	−6.70	to	2.69	1.000	−2.02	3.01	−7.91	to	3.87	1.000
	Initiating treatment						**Noninitiating treatment**					
MMF vs CYC	−1.46	1.11	−3.63	to	0.71	.758	−1.43	1.73	−4.82	to	1.97	1.000
CYC vs TCZ	4.50	2.44	−0.29	to	9.29	.268	2.95	2.54	−2.03	to	7.94	.992
MMF vs TCZ	3.04	2.39	−1.65	to	7.74	.824	1.53	2.24	−2.86	to	5.91	1.000
CYC vs RTX	1.94	1.32	−0.65	to	4.53	.572	2.07	2.07	−2.00	to	6.13	1.000
MMF vs RTX	0.48	1.23	−1.92	to	2.89	.992	0.64	1.69	−2.66	to	3.95	1.000
TCZ vs RTX	−2.56	2.50	−7.46	to	2.34	1.000	−0.89	2.51	−5.81	to	4.04	1.000
	Current nonsmoker						**Current smoker**					
MMF vs CYC	−1.43	1.16	−3.70	to	0.84	.866	−1.47	1.53	−4.47	to	1.52	1.000
CYC vs TCZ	4.88	2.48	0.01	to	9.74	.203	2.26	2.37	−2.38	to	6.90	1.000
MMF vs TCZ	3.44	2.40	−1.27	to	8.15	.619	0.79	2.20	−3.53	to	5.11	1.000
CYC vs RTX	1.24	1.35	−1.41	to	3.88	1.000	4.48	2.44	−0.30	to	9.26	.271
MMF vs RTX	−0.20	1.20	−2.56	to	2.16	1.000	3.01	2.28	−1.47	to	7.48	.758
TCZ vs RTX	−3.64	2.50	−8.54	to	1.27	.593	2.22	2.91	−3.49	to	7.92	1.000

* Calculated by average treatment effects of the propensity score-based inverse probability of treatment weighting model.
CYC, cyclophosphamide; IS, immunosuppressant; MMF, mycophenolate mofetil; RTX, rituximab; TCZ, tocilizumab.

impact decision-making and effectiveness, emphasizing the need for contextual interpretation. Lastly, observational effectiveness studies cannot provide the same level of evidence as prospective, randomized, controlled clinical trials. However, head-to-head comparison trials with 4 different agents are unlikely to be feasible in an orphan disease such as SSc-ILD, and effectiveness trials provide important information under these circumstances.

In conclusion, in this first large real-world study, the effectiveness of FVC change in SSc-ILD patients was not statistically different between TCZ, RTX, MMF, and CYC. Head-to-head comparisons from RCTs and a large safety profile analysis are needed for SSc-ILD.

Competing interests

QY: received speaker fees from Boehringer Ingelheim. ME: congress support from Janssen and AstraZeneca outside of the submitted work. CB: consultant for Boehringer Ingelheim; grant/research support from Gruppo Italiano Lotta alla Sclerodermia (GILS), European Scleroderma Trials and Research

Group (EUSTAR), Foundation for Research in Rheumatology (FOREUM), Scleroderma Clinical Trials Consortium (SCTC), Scleroderma Research Foundation (SRF), and Novartis Foundation for Medical-Biological Research. Educational grants from AbbVie and the Wellcome Trust. Congress support from Boehringer Ingelheim. CM: received speaker and/or consultancy fees from Janssen-Cilag AG, Boehringer Ingelheim, MED Talks Switzerland, Medbase, Mepha, Novartis, and PlayToKnow AG, and congress support from Boehringer Ingelheim, Novartis, and Roche. RD: speakers bureau: Actelion; consultancy: Boehringer Ingelheim; grant/research support from Pfizer, Actelion, and Iten-Kohaut Foundation; congress/workshop participation support: Amgen (2022) and Otsuka (2023), unrelated to the present work. OD has/had consultancy relationship with and/or has received research funding from and/or has served as a speaker for the following companies in the area of potential treatments for SSc and its complications in the last 3 calendar years: 4P-Pharma, Abbvie, Acceleron, Alcimed, Altavant, Amgen, AnaMar, Argenx, Arxx, AstraZeneca, Blade, Bayer, Boehringer Ingelheim, Corbus, CSL Behring, Galderma, Galapagos, Glenmark, Gossamer, Horizon, Janssen, Kymera, Lupin, Medscape, Merck, Miltenyi Biotec, Mitsubishi Tanabe, Novartis, Orion, Prometheus, Redxpharma, Roivant, Topadur, and UCB. Patent issued "mir-29 for the treatment of systemic sclerosis" (US8247389, EP2331143). Cofounder of SCITU AG. The other authors declare no conflict of interest.

Acknowledgements

The authors thank Ms Sun An for her kind support in statistics. The authors thank all health professionals and patients who participated in the registers.

Contributors

All the authors have provided substantial contributions to the conception or design of the work, the acquisition of the data, and the interpretation of data. QY performed the statistical analysis. QY wrote the first draft. All the other authors participated in the final drafting of the work or revising it critically for important intellectual content. All authors contributed to the final approval of the version published. QY and OD accept full responsibility for the work and the conduct of the study, had access to the data, and controlled the decision to publish. QY is the guarantor.

Funding

QY was supported by the China Scholarship Council and a training grant by Shanghai Leadingtac Pharmaceuticals to the University of Zurich.

Patient consent for publication

Not applicable.

Ethics approval

Local ethics committees (when required according to local legislation) of the respective EUSTAR centres approved the collection of data. The analysis was approved by the ethics committee of the Canton of Zurich (BASEC Nr. 2018-02165).

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

Data are available upon reasonable request. Anonymized data can be shared based on EUSTAR approval and regulations.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.014.

REFERENCES

[1] Allanore Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP, et al. Systemic sclerosis. Nat Rev Dis Primers 2015;1:15002. doi: 10.1038/nrdp.2015.2.
[2] Ranque B, Mouthon L. Geoepidemiology of systemic sclerosis. Autoimmun Rev 2010;9(5):A311–8. doi: 10.1016/j.autrev.2009.11.003.
[3] Elhai M, Meune C, Boubaya M, Avouac J, Hachulla E, Balbir-Gurman A, et al. Mapping and predicting mortality from systemic sclerosis. Ann Rheum Dis 2017;76(11):1897–905. doi: 10.1136/annrheumdis-2017-211448.
[4] Cottin V, Brown KK. Interstitial lung disease associated with systemic sclerosis (SSc-ILD). Respir Res 2019;20(1):13. doi: 10.1186/s12931-019-0980-7.
[5] Hoffmann-Vold AM, Fretheim H, Halse AK, Seip M, Bitter H, Wallenius M, et al. Tracking impact of interstitial lung disease in systemic sclerosis in a complete nationwide cohort. Am J Respir Crit Care Med 2019;200(10):1258–66. doi: 10.1164/rccm.201903-0486OC.
[6] Yin H, Li R, Lu L, Yan Q. Understanding fibrosis in systemic sclerosis: novel and emerging treatment approaches. Curr Rheumatol Rep 2020;22(11):77. doi: 10.1007/s11926-020-00953-0.
[7] Bournia VK, Fragoulis GE, Mitrou P, Mathioudakis K, Tsolakidis A, Konstantonis G, et al. All-cause mortality in systemic rheumatic diseases under treatment compared with the general population, 2015-2019. RMD Open 2021;7(3):e001694. doi: 10.1136/rmdopen-2021-001694.
[8] Galdo FD, Lescoat A, Conaghan PG, Ananyeva P, Balbir-Gurman A, Bertoldo E, et al. OP0234 2023 update of EULAR recommendations for the treatment of systemic sclerosis. Ann Rheum Dis 2023;82(Suppl 1):154S–5S. doi: 10.1136/annrheumdis-2023-eular.1383.
[9] Pope JE, Denton CP, Johnson SR, Fernandez-Codina A, Hudson M, Nevskaya T. State-of-the-art evidence in the treatment of systemic sclerosis. Nat Rev Rheumatol 2023;19(4):212–26. doi: 10.1038/s41584-023-00909-5.
[10] Tashkin DP, Elashoff R, Clements PJ, Goldin J, Roth MD, Furst DE, et al. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354(25):2655–66. doi: 10.1056/NEJMoa055120.
[11] Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. Lancet Respir Med 2016;4(9):708–19. doi: 10.1016/s2213-2600(16)30152-7.
[12] Khanna D, Lin CJF, Furst DE, Goldin J, Kim G, Kuwana M, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Respir Med 2020;8(10):963–74. doi: 10.1016/S2213-2600(20)30318-0.
[13] Khanna D, Denton CP, Lin CJF, van Laar JM, Frech TM, Anderson ME, et al. Safety and efficacy of subcutaneous tocilizumab in systemic sclerosis: results from the open-label period of a phase II randomised controlled trial (faSSciinate). Ann Rheum Dis 2018;77(2):212–20. doi: 10.1136/annrheumdis-2017-211682.
[14] Ebata S, Yoshizaki A, Oba K, Kashiwabara K, Ueda K, Uemura Y, et al. Safety and efficacy of rituximab in systemic sclerosis (DESIRES): open-label extension of a double-blind, investigators-initiated, randomised, placebo-controlled trial. Lancet Rheumatol 2022;4(8):e546–e55. doi: 10.1016/S2665-9913(22)00131-X.
[15] Hoffmann-Vold AM, Brunborg C, Airò P, Ananyeva LP, Czirják L, Guiducci S, et al. Cohort enrichment strategies for progressive interstitial lung disease in systemic sclerosis from European Scleroderma Trials and Research. Chest 2023;163(3):586–98. doi: 10.1016/j.chest.2022.09.044.
[16] Maher TM, Tudor VA, Saunders P, Gibbons MA, Fletcher SV, Denton CP, et al. Rituximab versus intravenous cyclophosphamide in patients with connective tissue disease-associated interstitial lung disease in the UK (RECITAL): a

double-blind, double-dummy, randomised, controlled, phase 2b trial. Lancet Respir Med 2023;11(1):45–54. doi: 10.1016/S2213-2600(22)00359-9.
[17] Sircar G, Goswami RP, Sircar D, Ghosh A, Ghosh P. Intravenous cyclophosphamide vs rituximab for the treatment of early diffuse scleroderma lung disease: open label, randomized, controlled trial. Rheumatology (Oxford) 2018;57(12):2106–13. doi: 10.1093/rheumatology/key213.
[18] Jordan S, Distler JH, Maurer B, Huscher D, van Laar JM, Allanore Y, et al. Effects and safety of rituximab in systemic sclerosis: an analysis from the European Scleroderma Trial and Research (EUSTAR) group. Ann Rheum Dis 2015;74(6):1188–94. doi: 10.1136/annrheumdis-2013-204522.
[19] Meier FM, Frommer KW, Dinser R, Walker UA, Czirjak L, Denton CP, et al. Update on the profile of the EUSTAR cohort: an analysis of the EULAR Scleroderma Trials and Research group database. Ann Rheum Dis 2012;71(8):1355–60. doi: 10.1136/annrheumdis-2011-200742.
[20] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335(7624):806–8. doi: 10.1136/bmj.39335.541782.AD.
[21] van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 Classification criteria for systemic sclerosis: an American College of Rheumatology/European League against rheumatism collaborative initiative. Ann Rheum Dis 2013;72(11):1747–55. doi: 10.1136/annrheumdis-2013-204424.
[22] Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Arthritis Rheum 1980;23(5):581–90. doi: 10.1002/art.1780230510.
[23] Hoffmann-Vold AM, Allanore Y, Alves M, Brunborg C, Airó P, Ananieva LP, et al. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. Ann Rheum Dis 2021;80(2):219–27. doi: 10.1136/annrheumdis-2020-217455.
[24] Maurer B, Graf N, Michel BA, Müller-Ladner U, Czirják L, Denton CP, et al. Prediction of worsening of skin fibrosis in patients with diffuse cutaneous systemic sclerosis using the EUSTAR database. Ann Rheum Dis 2015;74(6):1124–31. doi: 10.1136/annrheumdis-2014-205226.
[25] Kafaja S, Clements PJ, Wilhalme H, Tseng CH, Furst DE, Kim GH, et al. Reliability and minimal clinically important differences of forced vital capacity: results from the Scleroderma Lung Studies (SLS-I and SLS-II). Am J Respir Crit Care Med 2018;197(5):644–52. doi: 10.1164/rccm.201709-1845OC.
[26] Khanna D, Clements PJ, Volkmann ER, Wilhalme H, Tseng CH, Furst DE, et al. Minimal clinically important differences for the modified Rodnan Skin Score: results from the Scleroderma Lung Studies (SLS-I and SLS-II). Arthritis Res Ther 2019;21(1):23. doi: 10.1186/s13075-019-1809-y.
[27] Courvoisier DS, Lauper K, Kedra J, de Wit M, Fautrel B, Frisell T, et al. EULAR points to consider when analysing and reporting comparative effectiveness research using observational data in rheumatology. Ann Rheum Dis 2022;81(6):780–5. doi: 10.1136/annrheumdis-2021-221307.
[28] McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med 2013;32(19):3388–414. doi: 10.1002/sim.5753.
[29] llanore Y, Wung P, Soubrane C, Esperet C, Marrache F, Bejuit R, et al. A randomised, double-blind, placebo-controlled, 24-week, phase II, proof-of-concept study of romilkimab (SAR156597) in early diffuse cutaneous systemic sclerosis. Ann Rheum Dis 2020;79(12):1600–7. doi: 10.1136/annrheumdis-2020-218447.
[30] Distler O, Highland KB, Gahlemann M, Azuma A, Fischer A, Mayes MD, et al. Nintedanib for systemic sclerosis-associated interstitial lung disease. N Engl J Med 2019;380(26):2518–28. doi: 10.1056/NEJMoa1903076.
[31] Highland KB, Distler O, Kuwana M, Allanore Y, Assassi S, Azuma A, et al. Efficacy and safety of nintedanib in patients with systemic sclerosis-associated interstitial lung disease treated with mycophenolate: a subgroup analysis of the SENSCIS trial. Lancet Respir Med 2021;9(1):96–106. doi: 10.1016/s2213-2600(20)30330-1.
[32] Johnson SR, Bernstein EJ, Bolster MB, Chung JH, Danoff SK, George MD, et al. 2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) Guideline for the treatment of interstitial lung disease in people with systemic autoimmune rheumatic diseases. Arthritis Rheumatol 2024;76(8):1182–200. doi: 10.1002/art.42861.
[33] Jee AS, Stewart I, Youssef P, Adelstein S, Lai D, Hua S, et al. A Composite serum biomarker index for the diagnosis of systemic sclerosis-associated interstitial lung disease: a multicenter, observational cohort study. Arthritis Rheumatol 2023;75(8):1424–33. doi: 10.1002/art.42491.

Ann Rheum Dis 84 (2025) 632–642

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Osteoarthritis

Association of walking with incident knee osteoarthritis: a prospective cohort study using data from the UK Biobank

Hongyi He[1,†], Yilun Wang[1,2,3,†], David Hunter[4,5], Yuqing Zhang[6,7], Weiya Zhang[8,9], Michael Doherty[8,9], Daniel K. White[10], Jie Wei[1,2,3,11,12], Chao Zeng[1,2,3,11,13,*], Guanghua Lei[1,2,3,13]

[1] *Department of Orthopaedics, Xiangya Hospital, Central South University, Changsha, China*
[2] *Key Laboratory of Aging-related Bone and Joint Diseases Prevention and Treatment, Ministry of Education, Xiangya Hospital, Central South University, Changsha, China*
[3] *Hunan Key Laboratory of Joint Degeneration and Injury, Xiangya Hospital, Central South University, Changsha, China*
[4] *Sydney Musculoskeletal Health, Kolling Institute, University of Sydney, Sydney, NSW, Australia*
[5] *Rheumatology Department, Royal North Shore Hospital, Sydney, NSW, Australia*
[6] *Division of Rheumatology, Allergy, and Immunology, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA*
[7] *The Mongan Institute, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA*
[8] *Academic Rheumatology, School of Medicine, University of Nottingham, Nottingham, UK*
[9] *Pain Centre Versus Arthritis, Nottingham, UK*
[10] *Department of Physical Therapy, University of Delaware, Newark, Delaware, USA*
[11] *Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China*
[12] *FuRong Laboratory, Changsha, Hunan, China*
[13] *National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China*

A R T I C L E I N F O

Article history:
Received 21 August 2024
Received in revised form 7 November 2024
Accepted 2 December 2024

A B S T R A C T

Objectives: Walking offers numerous health benefits, yet its association with symptomatic knee osteoarthritis (SKOA) remains uncertain. The conflation of purposeful walking and unintentional walking could contribute to this uncertainty. Therefore, we aimed to examine the association between daily purposeful walking steps, daily unintentional walking steps, and total daily steps with incident SKOA.

Methods: We conducted a population-based cohort study using data from the UK Biobank. Of the participants who had valid accelerometer data at baseline but no history of SKOA (identified through primary care or hospitalisation records), 89,969 were included. Hazard ratios (HRs) for categories of daily purposeful walking steps (cadence ≥60 steps/min), daily unintentional walking steps (cadence <60 steps/min), and total daily steps were estimated using a Cox proportional hazard model. The dose-response relationship was assessed using restricted cubic spline regression.

Results: During a mean follow-up of 6.85 (SD, 1.14) years, 2711 participants developed SKOA. For daily purposeful walking steps, HRs (95% CIs) of incident SKOA were 0.84 (0.76-0.92), 0.81 (0.71-0.90), and 0.74 (0.64-0.85) for those walking 4000 to 5999, 6000 to 7999, and

*Correspondence to Prof Chao Zeng.
E-mail addresses: weij1988@csu.edu.cn (J. Wei), zengchao@csu.edu.cn (C. Zeng), lei_guanghua@csu.edu.cn (G. Lei).
Handling editor Josef S. Smolen.
† HH and YW contributed equally.

https://doi.org/10.1016/j.ard.2025.01.019

≥8000 steps/d, respectively, compared with those walking <4000 steps/d. Restricted cubic spline regression analysis revealed a declining curve with a potential threshold near 8000 steps/d. Conversely, more daily unintentional walking steps were significantly associated with higher incident SKOA. No significant association was observed for total daily steps.
Conclusions: More daily purposeful walking steps were associated with lower incident SKOA, suggesting that purposeful walking may promote musculoskeletal health and help inform future guidelines for knee osteoarthritis prevention.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Symptomatic knee osteoarthritis (SKOA) is a highly prevalent condition that causes significant disability and is becoming an even greater societal burden with ageing populations and rising obesity rates. However, no effective preventative strategy exists as modifying its known risk factors remains a significant challenge.
- Walking is an accessible, cost-effective, low- to moderate-impact physical activity that offers numerous health benefits, yet its association with SKOA remains uncertain. A key factor contributing to this uncertainty may be the failure to distinguish between purposeful walking and unintentional walking, which differ in both intensity of effort and mechanical loading.

WHAT THIS STUDY ADDS

- The incidence of SKOA decreased steadily with an increase in daily purposeful walking steps (defined as steps taken at a cadence of ≥60 steps/min) up to approximately 8000 steps. This suggests that targeting a minimum of 8000 purposeful steps/day may be beneficial, after which the benefits plateau.
- Daily unintentional walking steps (defined as steps taken at a cadence of <60 steps/min) were significantly associated with higher incident SKOA, while no significant association was observed between total daily steps and incident SKOA.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- Increasing the number of daily purposeful walking steps at a higher cadence may help prevent SKOA.

INTRODUCTION

Knee osteoarthritis (OA) is a highly prevalent condition, affecting approximately 364 to 654 million adults aged ≥40 years worldwide [1,2]. Pain from knee OA is a major cause of disability and the primary reason for patients to seek medical care [3]. With global trends of an ageing population and increasing obesity prevalence, symptomatic knee OA (SKOA) is projected to become an even greater societal burden [1,4]. Among the various constitutional and environmental risk factors identified for knee OA, the main modifiable factors include overweight or obesity, repetitive occupational or recreational microtrauma, and biomechanical factors such as alterations in gait mechanics. However, modifying these factors remains a significant challenge [4,5].

Walking is an accessible, cost-effective, low- to moderate-impact physical activity, making it a popular choice for maintaining physical activity levels [6]. It offers numerous health benefits, including reducing the risk of overweight and obesity, diabetes, cardiovascular disease-specific mortality, and all-cause mortality [7–10]. However, the association between walking and the risk of developing knee OA remains inconclusive. Some studies suggest that walking is beneficial for knee OA, contributing to symptom relief and structural improvements, such as pain relief, reductions in cartilage degeneration, and lower risk of knee or hip replacement [11–15]. However, other studies report no association [16–20] or even contradictory results regarding the impact of walking on knee OA [21]. A key limitation in previous studies is the failure to distinguish between 'unintentional walking' and 'purposeful walking'. Although closely related, unintentional walking and purposeful walking are distinct. Unintentional walking encompasses unintentional or spontaneous movements (eg, incidental or sporadic steps) such as stretching while standing or intentional movements such as a few steps to go to the bathroom, and all are at low effort with low cadences [22]. In contrast, purposeful walking is a coordinated, rhythmic movement with higher cadences and intensity and includes continuous walking for sustained periods, such as walking for exercise or commuting [22]. Combining purposeful walking with unintentional walking may lead to inaccurate assessments of their effect on knee OA, contributing to inconsistent findings. Distinguishing between these step accumulation patterns is critical for accurately understanding the relation of daily walking to the risk of knee OA.

To address this knowledge gap, we conducted a large-scale population-based cohort study using data from the UK Biobank. We differentiated between purposeful walking, defined as steps taken at a cadence ≥60 steps/min, and unintentional walking, defined as steps taken at a cadence <60 steps/min. This approach allowed us to examine the associations of both purposeful and unintentional walking with the incidence of SKOA.

METHODS

Data source and study sample

The UK Biobank, established between 2006 and 2010, is a prospective cohort study comprising 502,412 participants aged 37 to 73 years across the UK [23]. The study was approved by the National Health Service (NHS) and the National Research Ethics Service (reference 11/NW/0382). All participants provided written informed consent. Initial data collection during recruitment involved questionnaires and physical assessments, supplemented by ongoing follow-up through linkage to national health-related datasets for outcome monitoring. Between 2013 and 2015, a subset of individuals (n = 103,624) who consented to additional monitoring were instructed to continuously wear an Axivity AX3 accelerometer on their dominant wrist for 24 hours daily over 7 consecutive days. For the current analysis, participants had to have worn the accelerometer for at least 16 hours daily for ≥3 days [24]. Individuals with a pre-existing diagnosis of SKOA at baseline and those without cadence data were excluded from the study.

Assessment of walking

Baseline steps were derived from raw acceleration data (collected between 2013 and 2015) using the validated Verisene

activity algorithm designed for wrist accelerometers [25]. Total daily steps were the average number of steps per day across all valid days. Purposeful walking was defined as the step accumulation pattern with a cadence of ≥60 steps/min. This cadence-based definition captures both the quantity and speed of walking, which is particularly relevant for individuals engaging in structured and planned physical activities [22]. Daily purposeful walking steps were determined by averaging the daily purposeful walking step counts on valid days. Additionally, unintentional walking was measured as the step accumulation pattern with a cadence of <60 steps/min. Detailed definitions of these measures are provided in the Supplementary Material.

Assessment of incident knee OA

The diagnosis of SKOA was determined using the 'first occurrence of health outcomes' field in the UK Biobank database, identified by the International Statistical Classification of Diseases and Related Health Problems Tenth Revision (ICD-10) code M17 [26]. These diagnoses were based on data linked from the primary care records, hospital admission records (95.1%), or self-reported medical conditions (4.9%) [27]. Incident SKOA was defined as the diagnosis of SKOA occurring at any time after the index date, which was the date of baseline accelerometer data collection.

Assessment of other baseline characteristics

Baseline information was collected during recruitment (from 2006 to 2010) or ascertained using dates corresponding to hospital admission records. Age, sex, and Townsend deprivation index score were obtained from local NHS Primary Care Trust registries. Baseline body mass index (BMI) was calculated based on manually measured height and weight. Information on education, ethnicity, alcohol drinking status, smoking status, nonsteroidal anti-inflammatory drug (NSAIDs) use, and health status was collected via a touchscreen questionnaire. Knee injury diagnosis was ascertained based on the ICD-10 codes (S81-S89). Participants were defined with a history of knee injury if their diagnosis occurred at any time before the date of baseline accelerometer data collection. Frailty status was determined using 5 frailty phenotype indicators: grip strength, slow walking speed, exhaustion, recent weight loss, and physical activity [28]. Accelerometer wear time (valid hours and days) and meteorologic seasons were extracted from accelerometer data [16].

Statistical analysis

We calculated the quartile values for total daily steps, daily purposeful walking steps, and daily unintentional walking steps in the UK Biobank population [8,9]. Based on these quartiles, we selected rounded integer values as cutoffs to classify activity levels into 4 groups to aid future clinical recommendations [8,9,29]. Specifically, the cutoff values were 8000, 10,000, and 12,000 steps/d for total daily steps; 4000, 6000, and 8000 steps/d for daily purposeful walking steps; and 4000, 5000, and 6000 steps/d for daily unintentional walking steps. Common clinical benchmarks, with <4000 steps/d indicating inactivity and >8000 steps/d indicating high activity, were included in our categorisation [29]. Participants were followed from the study baseline (index date) until the first diagnosis of SKOA, death, loss to follow-up, or the end of follow-up (December 2021), whichever occurred first. We compared baseline characteristics across 4 categories of exposure groups using standardised mean difference, with a threshold of ≤0.1 indicating minimal imbalance [30]. We estimated the hazard ratios (HRs) and the corresponding 95% CIs for incident SKOA in the second, third, and fourth groups of exposures compared with the corresponding reference group using a Cox proportional hazards model with age as the timescale. The proportional hazard assumption (that HRs for different levels of exposures are constant over time) was tested by evaluating the statistical significance of interactions between age and exposures.

We used a causal diagram (Supplementary Figure S1) to guide our multivariable-adjusted analyses. Specifically, in evaluating the association between total daily steps and SKOA, we adjusted for sex, education, ethnicity, Townsend deprivation index score, alcohol drinking status, smoking status, health status, knee injury history, frailty, meteorologic season, and accelerometer wear time in the primary analyses (model 1). However, when examining the association of daily purposeful walking steps and daily unintentional walking steps with SKOA, we additionally adjusted for each type of walking in the analysis to account for potential confounding effects (model 2). Considering that previous studies have indicated that walking can affect BMI and medication use (eg, NSAID usage) [31,32], these variables could act as potential mediators in the association between walking and incident SKOA. However, they may also serve as confounders. Therefore, we further adjusted for these variables (model 3). Death was treated as a competing event and adjusted for in all the above models using the Fine-Gray method [33]. Tests for trends across exposure groups were conducted by adding each group's median value to the regression model. To better understand the relationship between walking steps and the risk of SKOA, we used restricted cubic splines to smooth the dose-response curve [34]. Subgroup analyses explored whether the association (if one exists) between exposures and SKOA risk varied by age (<65 years old, ≥65 years old) and sex, given their potential influence on walking steps and SKOA risk. We tested interaction effects by including interaction terms (eg, sex × purposeful walking groups, or age category × purposeful walking groups) in the regression model.

We conducted 3 sensitivity analyses. First, to minimise the possibility of reverse causality, we implemented a 6-month and 12-month exposure lag time by excluding SKOA cases that occurred within 6 or 12 months after the index date to evaluate the robustness of our study findings. Second, we removed potential running steps from purposeful walking steps using a cutoff value of 120 steps/min [22]. Third, we calculated the E-value to quantify the minimum strength of association that an unmeasured confounder would need to nullify the associations observed in analyses [35].

All analyses were conducted using R version 4.2.2, and a 2-sided P value ≤ .05 was considered statistically significant for all tests.

Patient and public involvement

This research did not involve input from patients or the public in its planning and dissemination. However, steps have been outlined to share findings with stakeholders. After publication, the study results will be disseminated through the authors' institutions' media centre, with formal press statements, and dissemination efforts will also utilise networks and social media platforms to maximise outreach and engagement.

RESULTS

Among the 98,248 participants with accelerometer data from the UK Biobank, we excluded 3825 individuals with invalid accelerometer data, 4453 with prevalent SKOA, and 1 with missing walking steps data. Among the remaining 89,969 participants, the mean age was 56.0 years, 56.5% were women, and the average number of total daily steps was 10,896 (Supplementary Figure S2). During a mean follow-up of 6.85 years (SD, 1.14), 2711 participants developed SKOA.

Association between total daily steps and incident SKOA

Table 1 presents the baseline characteristics of participants across the 4 groups categorised by total daily steps. Compared with those in the reference group (ie, <8000 total daily steps), participants in the higher total daily step groups were generally younger, had a higher proportion of women, lower BMI, better self-reported health status, and lower prevalence of frailty.

Compared with the reference group, the HRs and 95% CIs for incident SKOA in the second (8000-9999), third (10,000-11,999), and fourth groups (≥12,000 steps/d) were 0.87 (0.78-0.97), 0.75 (0.67-0.84), and 0.75 (0.68-0.84), respectively (Table 2). However, no significant association between total daily steps and incident SKOA was observed after adjusting for major confounders or when excluding participants who developed SKOA during the first 6 or 12 months of follow-up (Table 2 and Figure 1).

Association between daily purposeful walking steps and incident SKOA

Compared with those in the reference group (ie, <4000 daily purposeful walking steps), participants in the higher daily purposeful walking step groups were generally younger, had lower BMI, higher education levels, better self-reported health status, as well as lower prevalence of frailty and history of knee injury (Table 3).

As shown in Table 4, the incidence rates of SKOA were 6.16, 4.10, 3.53, and 3.03 per 1000 person-years in the reference (<4000), second (4000-5999), third (6000-7999), and fourth groups (≥8000 steps/d) of daily purposeful waking steps, respectively. The HRs and 95% CIs for incident SKOA in the second, third, and fourth groups were 0.79 (0.72-0.87), 0.73 (0.66-0.82), and 0.65 (0.57-0.75), respectively, compared with the reference group. Given the positive correlation between purposeful and unintentional walking steps (Spearman's correlation coefficient = 0.45, $P < .001$), the associations between daily purposeful walking steps and incident SKOA strengthened after adjusting for daily unintentional walking steps (model 2). Further adjustments for BMI and NSAID use (model 3), excluding participants who developed SKOA within the first 6 or 12 months of follow-up, and removing potential running steps all attenuated the association; however, a robust dose-response relationship persisted (all P for trend < .001). As depicted in the Figure 1, restricted cubic spline regression analysis revealed an inverse association, with HRs for incident SKOA significantly

Table 1
Baseline characteristics of participants across the 4 groups of total daily steps

Characteristics	Groups according to total daily steps (steps/d)				SMD
	Reference (<7999)	Second (8000-9999)	Third (10,000-11,999)	Fourth (≥12,000)	
Participants, N	17,708	19,789	21,485	30,987	
Age, y, mean (SD)	58.27 (7.60)	56.40 (7.87)	55.57 (7.80)	54.69 (7.65)	0.250
Men, n (%)	9182 (51.9)	8965 (45.3)	9056 (42.2)	11,936 (38.5)	0.145
BMI, n (%)					0.347
Normal	4544 (25.8)	7037 (35.6)	9058 (42.2)	15,657 (50.6)	
Overweight	7424 (42.1)	8536 (43.2)	9076 (42.3)	11,888 (38.4)	
Obese	5655 (32.1)	4178 (21.2)	3314 (15.5)	3408 (11.0)	
Deprivation index, mean (SD)	−1.66 (2.88)	−1.85 (2.76)	−1.80 (2.77)	−1.65 (2.85)	0.044
White, n (%)	17,153 (97.3)	19,085 (96.8)	20,751 (96.9)	29,887 (96.8)	0.016
College, n (%)	6861 (39.3)	8512 (43.5)	9833 (46.1)	13,967 (45.5)	0.076
Alcohol, n (%)					0.046
Never	625 (3.5)	554 (2.8)	574 (2.7)	831 (2.7)	
Previous	645 (3.6)	514 (2.6)	516 (2.4)	780 (2.5)	
Current	16,420 (92.8)	18,706 (94.6)	20,380 (94.9)	29,345 (94.8)	
Smoking, n (%)					0.088
Never	9325 (52.8)	11,366 (57.6)	12,560 (58.6)	18,319 (59.3)	
Previous	6644 (37.6)	6986 (35.4)	7577 (35.3)	10,716 (34.7)	
Current	1683 (9.5)	1396 (7.1)	1306 (6.1)	1861 (6.0)	
NSAIDs use, n (%)	4882 (28.0)	5260 (26.9)	5543 (26.1)	8028 (26.2)	0.024
Health status, n (%)					0.262
Excellent	2560 (14.5)	4051 (20.5)	5136 (23.9)	8194 (26.5)	
Good	9961 (56.5)	12,036 (61.0)	13,100 (61.1)	18,899 (61.1)	
Fair	4087 (23.2)	3189 (16.2)	2873 (13.4)	3496 (11.3)	
Poor	1031 (5.8)	461 (2.3)	338 (1.6)	341 (1.1)	
Knee injury history, n (%)	305 (1.7)	327 (1.7)	321 (1.5)	431 (1.4)	0.016
Frailty, n (%)	787 (4.4)	343 (1.7)	207 (1.0)	184 (0.6)	0.139
Hours per valid day, mean (SD)	23.81 (0.52)	23.82 (0.55)	23.81 (0.58)	23.80 (0.62)	0.011
Valid day, mean (SD)	5.79 (0.61)	5.80 (0.58)	5.80 (0.59)	5.77 (0.63)	0.028
Invalid day, mean (SD)	1.17 (0.55)	1.15 (0.51)	1.16 (0.53)	1.18 (0.57)	0.032
Meteorologic seasons, n (%)					0.111
Spring	3493 (19.7)	4017 (20.3)	4572 (21.3)	7054 (22.8)	
Summer	4452 (25.1)	5150 (26.0)	5905 (27.5)	9454 (30.5)	
Autumn	5045 (28.5)	5718 (28.9)	6153 (28.6)	8583 (27.7)	
Winter	4718 (26.6)	4904 (24.8)	4855 (22.6)	5896 (19.0)	

BMI, body mass index; NSAID, nonsteroidal anti-inflammatory drugs; SMD, standardised mean difference.

Table 2
Associations of total daily steps with incident symptomatic knee osteoarthritis

Total daily steps	Reference group (≤7999 steps/d)	Second group (8000-9999 steps/d)	Third group (10,000-11,999 steps/d)	Fourth group (≥12,000 steps/d)	*P* for trend
N	17,709	19,789	21,485	30,987	
Incident case, n	699	636	574	802	
Incident rate (per 1000 person-years)	5.90	4.70	3.87	3.74	
Follow-up time, y, mean	6.69	6.84	6.90	6.93	
Case diagnosis time, y, mean	2.97	3.27	3.29	3.33	
Crude model, HR (95% CI)	1.00 (reference)	0.87 (0.78, 0.97)	0.75 (0.67, 0.84)	0.75 (0.68, 0.84)	<.001
Model 1, HR* (95% CI)	1.00 (reference)	0.94 (0.84, 1.05)	0.83 (0.74, 0.94)	0.86 (0.77, 0.96)	.002
Model 2, HR† (95% CI)	1.00 (reference)	0.98 (0.88, 1.10)	0.90 (0.80, 1.02)	0.99 (0.88, 1.10)	.633
6-month lag, HR‡ (95% CI)	1.00 (reference)	1.03 (0.92, 1.16)	0.93 (0.82, 1.05)	1.05 (0.94, 1.18)	.532
12-month lag, HR§ (95% CI)	1.00 (reference)	1.05 (0.93, 1.19)	0.97 (0.85, 1.10)	1.10 (0.98, 1.24)	.159

* HRs are adjusted for age, sex, education, health status, deprivation, ethnicity, smoking status, alcohol drinking status, knee injury history, frailty, and accelerometer wear time.
† HRs are adjusted for body mass index and nonsteroidal anti-inflammatory drug usage on the basis of model 1.
‡ Analysis was performed by excluding the knee osteoarthritis cases that developed within 6 months after the index date and adjusted for covariates the same as model 2.
§ Analysis was performed by excluding the knee osteoarthritis cases that developed within 12 months after the index date and adjusted for covariates the same as model 2.HR, hazard ratio.

decreasing with increasing daily purposeful walking steps up to 8000, beyond which the estimate plateaued. Consistent inverse associations between daily purposeful walking steps and incident SKOA were observed across different age and sex subgroups. The association appeared to be stronger in women than in men (*P* for interaction = .008), while no significant interaction by age group was found (*P* for interaction = .931) (Supplementary Figure S3). E-values were 2.02 (1.81-2.40), 2.35 (1.96-2.72), and 2.66 (2.21-3.18) for the second, third, and fourth groups of daily purposeful walking steps, respectively. This suggests that the relation of potential residual confounders to both daily purposeful walking steps and SKOA must be ≥2.02 to nullify the modestly association.

Association between daily unintentional walking steps and incident SKOA

The baseline characteristics of participants across the 4 groups categorised by daily unintentional walking steps were similar to those categorised by daily purposeful walking steps except for the education level (Table 5). Compared with those in the reference group (ie, <4000 daily unintentional walking steps), the HRs and 95% CIs for incident SKOA in the second (4000-4999), third (5000-5999), and fourth groups (≥6000 steps/d) were 1.24 (1.10-1.40), 1.42 (1.26-1.61), and 1.69 (1.49-1.96), respectively (Table 6). Restricted cubic spline regression also revealed a consistent significant positive dose-response relationship (Figure 1). Excluding participants who developed SKOA during the first 6 or 12 months of follow-up did not change the association materially. The association was stronger in men than in women (*P* for interaction = .014), while no significant interaction by age group was observed (*P* for interaction = 0.196) (Supplementary Figure S4). E-values were 1.79 (1.43-2.15), 2.19 (1.83-2.60), and 2.77 (2.34-3.27) in the second, third, and fourth groups of daily unintentional walking steps, respectively, indicating that the relation of potential residual confounders to both daily unintentional walking steps and SKOA must be ≥1.79 to nullify the modestly association.

No evidence of a violation of the proportional hazard assumptions (no interaction between age and exposures) was found in any of these analyses (all $P > .05$).

DISCUSSION

In this large population-based cohort study conducted in the UK, purposeful walking (ie, steps taken at a cadence of ≥60 steps/min) of 8000 steps or more daily was associated with a 26% lower risk of SKOA than walking <4000 steps. This association was more pronounced in women than in men. The incidence of SKOA decreased steadily with increasing daily purposeful walking steps up to approximately 8000 steps, beyond which the benefit plateaued. In contrast, unintentional walking appeared to have a potentially detrimental effect on incident SKOA, and no significant association was observed between total daily steps and incident SKOA. These findings suggest that combining higher daily purposeful walking steps with increased cadence provides the most substantial benefit in reducing the risk of SKOA.

Comparison with previous studies

To date, only a few studies have examined the association between daily steps and the development of knee OA, and the results have been inconclusive. Doré et al [21] found that total daily steps ≥10,000 were associated with the worsening of pre-existing knee structural lesions (bone marrow lesions, meniscal pathology, cartilage defects, and reduced cartilage volume) assessed by magnetic resonance imaging over 3 years in 405 community-derived adults. In contrast, several studies have indicated potential protective effects of higher daily steps on cartilage damage, osteophyte formation, and functional limitations, with incident functional limitation potentially serving as a proxy for incident OA [11,13,14,36,37]. Other studies have found no significant association between daily steps and OA-related outcomes, such as knee pain, knee replacement, and structural changes on imaging, particularly in individuals at higher risk of knee OA or with mild knee OA [12,16,17]. To our knowledge, no previous study has specifically distinguished between purposeful and unintentional walking steps. In the current study, we made a clear distinction between these 2 types of stepping activity, which differ in both intensity of effort and mechanical loading. This distinction allowed us to investigate the unique contributions of purposeful walking to incident SKOA in the general population. Our findings suggested that faster purposeful walking (ie, steps taken with a cadence of ≥60 steps/min) may confer greater benefits in reducing the risk of SKOA.

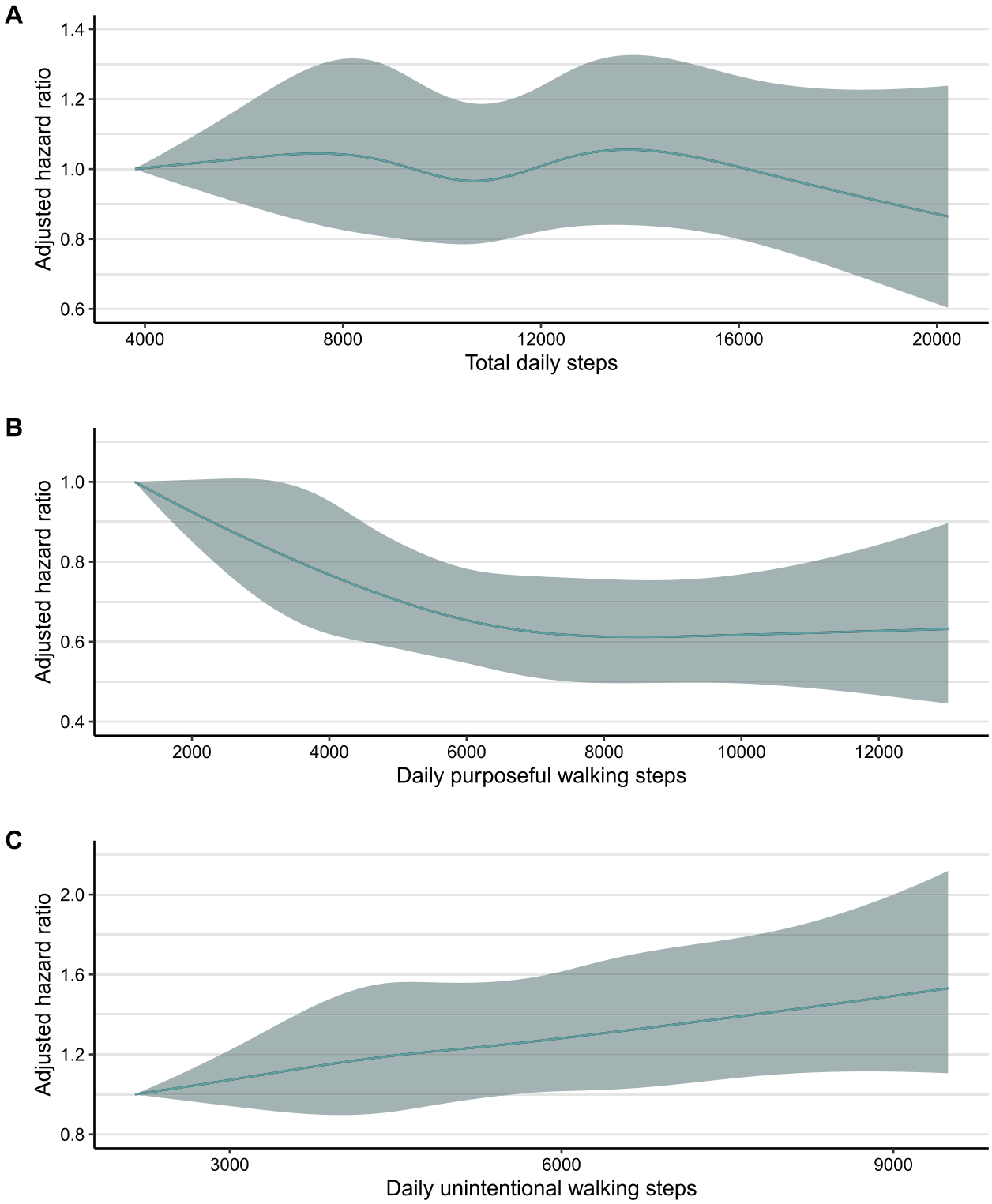


Figure. Associations of total daily steps, daily purposeful walking steps, and daily unintentional walking steps with incident symptomatic knee osteoarthritis. Association between total daily steps (A), daily purposeful walking steps (B), and daily unintentional walking steps (C) with incident symptomatic knee osteoarthritis. Data are shown as hazard ratios with 95% CIs. Hazard ratios are adjusted for sex, body mass index, education, ethnicity, Townsend deprivation index score, alcohol drinking status, smoking status, health status, nonsteroidal anti-inflammatory drug usage, knee injury history, frailty status, meteorologic season, and accelerometer wear time. The association of daily purposeful walking steps and daily unintentional walking steps with symptomatic knee osteoarthritis were mutually adjusted for each type of walking.

Strengths and limitations

Several strengths of our study are worth noting. First, we used accelerometers to objectively and prospectively collect data on walking volume and intensity. This approach provides a more accurate depiction of individuals' actual purposeful and unintentional walking levels than studies relying on self-reported data, which are susceptible to recall bias. Second, we applied a cadence threshold of ≥60 steps/min to differentiate purposeful walking from unintentional walking [22]. This distinction is important because purposeful walking and unintentional walking involve different intensity levels [22], which may help clarify the inconsistencies in previous findings regarding the association between walking and knee OA. Furthermore, our study demonstrated a

Table 3
Baseline characteristics of participants across the 4 groups of daily purposeful walking steps

Characteristics	Groups according to daily purposeful walking steps (steps/d)				SMD
	Reference (≤3999)	Second (4000-5999)	Third (6000-7999)	Fourth (≥8000)	
Participants, N	25,573	31,874	19,882	12,640	
Age, y, mean (SD)	57.80 (7.72)	55.74 (7.88)	54.99 (7.71)	54.48 (7.47)	0.232
Men, n (%)	11,338 (44.3)	13,526 (42.4)	8511 (42.8)	5764 (45.6)	0.037
BMI, n (%)					0.303
Normal	7481 (29.4)	13,111 (41.2)	9317 (46.9)	6387 (50.6)	
Overweight	10,572 (41.5)	13,352 (42.0)	8089 (40.7)	4911 (38.9)	
Obese	7412 (29.1)	5362 (16.8)	2451 (12.3)	1330 (10.5)	
Deprivation index, mean (SD)	−1.79 (2.80)	−1.86 (2.73)	−1.68 (2.84)	−1.37 (2.99)	0.091
White, n (%)	24,686 (96.9)	30,777 (96.9)	19,209 (97.0)	12,204 (96.9)	0.003
College, n (%)	9481 (37.6)	14,004 (44.3)	9513 (48.3)	6175 (49.3)	0.132
Alcohol, n (%)					0.053
Never	940 (3.7)	842 (2.6)	488 (2.5)	314 (2.5)	
Previous	897 (3.5)	761 (2.4)	474 (2.4)	323 (2.6)	
Current	23,711 (92.8)	30,248 (95.0)	18,903 (95.2)	11,989 (95.0)	
Smoking, n (%)					0.077
Never	13,729 (53.8)	18,651 (58.6)	11,821 (59.6)	7369 (58.5)	
Previous	9512 (37.3)	11,095 (34.9)	6824 (34.4)	4492 (35.6)	
Current	2258 (8.9)	2061 (6.5)	1182 (6.0)	745 (5.9)	
NSAIDs use, n (%)	7384 (29.3)	8285 (26.3)	5036 (25.7)	3008 (24.1)	0.062
Health status, n (%)					0.237
Excellent	3982 (15.6)	7433 (23.1)	5148 (25.9)	3467 (27.5)	
Good	14,795 (58.1)	19,432 (61.1)	12,143 (61.2)	7626 (60.5)	
Fair	5439 (21.3)	4494 (14.1)	2327 (11.7)	1385 (11.0)	
Poor	1261 (4.9)	543 (1.7)	230 (1.2)	137 (1.1)	
Knee injury history, n (%)	468 (1.8)	466 (1.5)	283 (1.4)	167 (1.3)	0.021
Frailty, n (%)	1031 (4.0)	308 (1.0)	125 (0.6)	57 (0.5)	0.132
Hours per valid day, mean (SD)	23.80 (0.55)	23.81 (0.57)	23.81 (0.58)	23.81 (0.65)	0.009
Valid day, mean (SD)	5.79 (0.61)	5.80 (0.58)	5.79 (0.61)	5.76 (0.67)	0.037
Invalid day, mean (SD)	1.16 (0.54)	1.15 (0.52)	1.17 (0.55)	1.19 (0.60)	0.036
Meteorologic seasons, n (%)					0.127
Spring	4985 (19.5)	6715 (21.1)	4469 (22.5)	2967 (23.5)	
Summer	6475 (25.3)	8736 (27.4)	5828 (29.3)	3922 (31.0)	
Autumn	7355 (28.8)	9132 (28.7)	5520 (27.8)	3492 (27.6)	
Winter	6758 (26.4)	7291 (22.9)	4065 (20.4)	2259 (17.9)	

BMI, body mass index; NSAID, nonsteroidal anti-inflammatory drug; SMD, standardised mean difference.

Table 4
Associations of daily purposeful walking steps with incident symptomatic knee osteoarthritis

Daily purposeful walking steps	Reference group (≤3999 steps/d)	Second group (4000-5999 steps/d)	Third group (6000-7999 steps/d)	Fourth group (≥8000 steps/d)	*P* for trend
N	25,573	31,874	19,882	12,640	
Incident case, n	1059	900	486	266	
Incident rate (per 1000 person-years)	6.16	4.10	3.53	3.03	
Follow-up time, y, mean	6.72	6.88	6.93	6.94	
Case diagnosis time, y, mean	3.08	3.28	3.26	3.44	
Crude model, HR (95% CI)	1.00 (reference)	0.73 (0.67, 0.80)	0.65 (0.58, 0.73)	0.57 (0.50, 0.66)	<.001
Model 1, HR[a] (95% CI)	1.00 (reference)	0.79 (0.72, 0.87)	0.73 (0.66, 0.82)	0.65 (0.57, 0.75)	<.001
Model 2, HR[b] (95% CI)	1.00 (reference)	0.70 (0.63, 0.77)	0.62 (0.55, 0.70)	0.54 (0.47, 0.63)	<.001
Model 3, HR[c] (95% CI)	1.00 (reference)	0.73 (0.66, 0.80)	0.67 (0.60, 0.76)	0.61 (0.53, 0.70)	<.001
6-mo lag, HR[d] (95%CI)	1.00 (reference)	0.75 (0.68, 0.83)	0.71 (0.63, 0.81)	0.65 (0.55, 0.75)	<.001
12-mo lag, HR[e] (95% CI)	1.00 (reference)	0.76 (0.68, 0.84)	0.73 (0.64, 0.83)	0.66 (0.57, 0.78)	<.001
Removing running steps (95% CI)	1.00 (reference)	0.83 (0.75, 0.91)	0.80 (0.70, 0.91)	0.80 (0.67, 0.96)	<.001

HR, hazard ratio; CI, confidence interval.

[a] HRs are adjusted for sex, education, ethnicity, Townsend deprivation index score, alcohol drinking status, smoking status, health status, knee injury history, frailty status, meteorologic season, and accelerometer wear time.

[b] HRs are adjusted for the daily unintentional walking steps on the basis of model 1.

[c] HRs are adjusted for body mass index and nonsteroidal anti-inflammatory drug usage on the basis of model 2.

[d] Analysis was performed by excluding the knee osteoarthritis cases that developed within 6 months after the index date and adjusted for covariates the same as model 3.

[e] Analysis was performed by excluding the knee osteoarthritis cases that developed within 12 months after the index date and adjusted for covariates the same as model 3.HR, hazard ratio.

Table 5
Baseline characteristics of participants across the 4 groups of daily unintentional walking steps

Characteristics	Groups according to daily unintentional walking steps (steps/d)				SMD
	Reference (≤3999)	Second (4000-4999)	Third (5000-5999)	Fourth (≥6000)	
Participants, N	16,803	21,401	22,521	29,244	
Age, y, mean (SD)	57.67 (7.69)	56.46 (7.81)	55.68 (7.80)	54.89 (7.77)	0.196
Men, n (%)	10,145 (60.4)	10,571 (49.4)	9191 (40.8)	9232 (31.6)	0.327
BMI, n (%)					0.289
Normal	4668 (27.9)	7556 (35.4)	9434 (42.0)	14,638 (50.1)	
Overweight	7370 (44.0)	9232 (43.2)	9409 (41.8)	10,913 (37.4)	
Obese	4708 (28.1)	4559 (21.4)	3642 (16.2)	3646 (12.5)	
Deprivation index, mean (SD)	−1.46 (2.97)	−1.72 (2.83)	−1.82 (2.78)	−1.82 (2.74)	0.068
White, n (%)	16,275 (97.3)	20,705 (97.1)	21,799 (97.1)	28,097 (96.4)	0.027
College, n (%)	7662 (46.2)	9876 (46.6)	9863 (44.2)	11,772 (40.6)	0.067
Alcohol, n (%)					0.046
Never	525 (3.1)	545 (2.5)	561 (2.5)	953 (3.3)	
Previous	578 (3.4)	565 (2.6)	561 (2.5)	751 (2.6)	
Current	15,685 (93.4)	20,270 (94.8)	21,384 (95.0)	27,512 (94.2)	
Smoking, n (%)					0.063
Never	9223 (55.0)	12,123 (56.8)	13,000 (57.8)	17,224 (59.1)	
Previous	6057 (36.1)	7700 (36.1)	7990 (35.6)	10,176 (34.9)	
Current	1476 (8.8)	1521 (7.1)	1485 (6.6)	1764 (6.0)	
NSAIDs use, n (%)	4064 (24.6)	5476 (25.9)	6070 (27.3)	8103 (28.1)	0.045
Health status, n (%)					0.187
Excellent	2873 (17.2)	4513 (21.1)	5204 (23.1)	7351 (25.2)	
Good	9516 (56.9)	12,929 (60.6)	13,724 (61.0)	17,827 (61.1)	
Fair	3568 (21.3)	3362 (15.7)	3144 (14.0)	3571 (12.2)	
Poor	781 (4.7)	547 (2.6)	408 (1.8)	435 (1.5)	
Knee injury history, n (%)	268 (1.6)	333 (1.6)	363 (1.6)	420 (1.4)	0.008
Frailty, n (%)	545 (3.2)	374 (1.7)	313 (1.4)	289 (1.0)	0.085
Hours per valid day, mean (SD)	23.83 (0.50)	23.82 (0.54)	23.81 (0.58)	23.79 (0.64)	0.034
Valid day, mean (SD)	5.78 (0.62)	5.81 (0.58)	5.79 (0.59)	5.77 (0.63)	0.032
Invalid day, mean (SD)	1.17 (0.55)	1.15 (0.51)	1.16 (0.53)	1.18 (0.57)	0.032
Meteorologic seasons, n (%)					0.086
Spring	3343 (19.9)	4491 (21.0)	4850 (21.5)	6452 (22.1)	
Summer	4227 (25.2)	5707 (26.7)	6313 (28.0)	8714 (29.8)	
Autumn	4903 (29.2)	6153 (28.8)	6253 (27.8)	8190 (28.0)	
Winter	4330 (25.8)	5050 (23.6)	5105 (22.7)	5888 (20.1)	

BMI, body mass index; NSAID, nonsteroidal anti-inflammatory drug; SMD, standardised mean difference.

Table 6
Associations of daily unintentional walking steps with incident symptomatic knee osteoarthritis

Daily unintentional walking steps	Reference group (≤3999 steps/d)	Second group (4000-4999 steps/d)	Third group (5000-5999 steps/d)	Fourth group (≥6000 steps/d)	*P* for trend
N	16,804	21,401	22,521	29,244	
Incident case, n	506	654	682	869	
Incident rate (per 1000 person-years)	4.47	4.47	4.40	4.30	
Follow-up time, y, mean	6.70	6.84	6.89	6.91	
Case diagnosis time, y, mean	3.05	3.22	3.27	3.27	
Crude model, HR (95%CI)	1.00 (reference)	1.07 (0.95, 1.20)	1.10 (0.98, 1.23)	1.12 (1.00, 1.25)	.055
Model 1, HR* (95%CI)	1.00 (reference)	1.12 (1.00, 1.27)	1.17 (1.04, 1.32)	1.21 (1.08, 1.36)	.002
Model 2, HR† (95%CI)	1.00 (reference)	1.23 (1.09, 1.38)	1.37 (1.21, 1.56)	1.57 (1.38, 1.79)	<.001
Model 3, HR‡ (95%CI)	1.00 (reference)	1.24 (1.10, 1.40)	1.42 (1.26, 1.61)	1.69 (1.49, 1.93)	<.001
6-mo lag, HR§ (95%CI)	1.00 (reference)	1.27 (1.12, 1.45)	1.45 (1.27, 1.66)	1.71 (1.49, 1.96)	<.001
12-mo lag, HR‖ (95%CI)	1.00 (reference)	1.28 (1.12, 1.46)	1.45 (1.26, 1.67)	1.76 (1.53, 2.03)	<.001

* HRs are adjusted for sex, education, ethnicity, Townsend deprivation index score, alcohol drinking status, smoking status, health status, knee injury history, frailty status, meteorologic season, and accelerometer wear time.

† HRs are adjusted for daily purposeful walking steps on the basis of model 1.

‡ HRs are adjusted for body mass index and nonsteroidal anti-inflammatory drug usage on the basis of model 2.

§ Analysis was performed by excluding the knee osteoarthritis cases that developed within 6 months after the index date and adjusted for covariates the same as model 3.

‖ Analysis was performed by excluding the knee osteoarthritis cases that developed within 12 months after the index date and adjusted for covariates the same as model 3.HR, hazard ratio.

dose-response relationship between purposeful walking steps, defined by both volume (ie, step counts) and cadence and the incidence of SKOA. This empirical evidence supports the development of a more precise and practical strategy for SKOA prevention.

Our study has several limitations. First, the use of an accelerometer may have led to an increased walking activity among the participants [38], potentially resulting in higher average daily purposeful or unintentional walking steps compared with the general population and those who did not consent to wear an accelerometer. However, the absolute values of steps did not appear to impact the association of walking with incident SKOA. Second, we required participants to wear the accelerometer for ≥3 days with a minimum of 16 valid hours per day, which may introduce bias due to varying physical activities throughout the week. However, our analysis revealed no significant difference in valid days between weekdays and weekends ($P = .93$), suggesting similar walking-related activities on both weekdays and weekends. Additionally, missing data could also introduce some bias if it is not random. However, since the invalid days were evenly distributed across different groups, they are unlikely to bias our results. Third, reverse causation could complicate the interpretation of our results, as early symptoms of SKOA might have influenced changes in walking levels before diagnosis [39]. Nevertheless, our sensitivity analyses, which incorporated a 6-month and 12-month lag, remained consistent with our primary results, supporting the validity of our findings. Fourth, the lack of radiographic data for scoring the presence of structural knee OA may have introduced potential misclassification in the diagnosis of SKOA, particularly for data derived from self-reported medical condition codes. However, the interviews in the UK Biobank were conducted by trained nurses, and additional diagnostic sources were linked to hospital inpatient data [27]. Therefore, any misclassification is likely to be nondifferential and would bias the results towards the null. Fifth, observational studies cannot completely rule out the potential for residual confounding and reverse causation. Additionally, UK Biobank participants are generally healthier and more active, averaging 10,896 steps/d compared with 9124 steps/d among US adults [8], which may limit the generalizability of our findings to other populations [9,40]. Further studies in diverse settings and over a longer duration, with repeated measures, are necessary to validate our findings.

Possible explanations

The mechanics of dynamic equilibrium between interstitial fluid loss and recovery in cartilage may explain the differing effects of daily purposeful versus unintentional walking on SKOA. Cartilage, composed of approximately 60% to 85% interstitial fluid [41], relies on this balance to maintain thickness and function [42,43]. Local fluid that flows within cartilage facilitates essential solute transport and biomechanical responses [41,43]. Higher-frequency loading from purposeful walking promotes the transport of nutrients and growth factors within the cartilage, enhancing chondrogenesis of mesenchymal stem cells and supporting cartilage protection [44]. This loading also creates substantial interstitial fluid pressure, reducing strain (ie, deformation) on the cartilage matrix and protecting it from excessive external mechanical loads [42,45,46]. In contrast, lower-frequency loading from unintentional walking allows interstitial fluid to flow more freely, resulting in increased cartilage deformation and a heightened susceptibility to mechanical strain [45]. Therefore, the cyclical and repetitive nature of purposeful walking likely accounts for its protective effect on SKOA [14].

Clinical implications

Numerous studies have demonstrated the extensive benefits of regular physical activity for both physical and mental health [14,47,48]. Due to these benefits, the American College of Rheumatology recommends exercise programmes for knee OA management [49]. However, despite these recommendations, concerns persist about whether physical activity might elevate OA risk owing to inconsistencies in the literature. Therefore, clarifying the effect of walking, one of the most common and easily undertaken physical activities, on the development of OA is crucial for public health. Such clarification could help optimise exercise prescriptions for knee OA patients and enhance public health strategies.

Our study highlights the clear benefits of purposeful walking for reducing the incidence of SKOA. These findings offer actionable data for public health initiatives aimed at OA prevention by identifying approximately 8000 daily purposeful walking steps as a beneficial target. Thus, guidelines should emphasise both the goal of achieving 8000 purposeful walking steps daily and the importance of walking at a higher cadence to protect against SKOA. Because up to 12,000 purposeful walking steps show no detrimental effects, we advise against reducing purposeful walking due to SKOA concerns. Additionally, since our model does not address the impact of exceeding 13,000 purposeful steps daily, the effect beyond this point remains unknown. While we observed a potential detrimental effect of unintentional walking on SKOA, it is important to note that unintentional walking is an inherent part of daily life. We therefore encourage individuals to include more purposeful walking. The combination of purposeful walking and unintentional walking may explain the lack of a significant association between total daily steps and SKOA in real-world settings, highlighting the importance of distinguishing between purposeful and unintentional walking in research.

The different effects of purposeful and unintentional walking on SKOA risk between men and women are particularly noteworthy. Men may engage in more unintentional walking due to occupational factors, while women often participate in planned or structured exercises. This could explain why women appear to experience greater protective benefits from purposeful walking while being less affected by unintentional walking. Future research should explore the biological or behavioural reasons behind these differences. Such insights could inform tailored intervention strategies and encourage revision of exercise guidelines for OA prevention to reflect sex-specific differences.

Furthermore, walking behaviours could serve as proxies for other exposures, such as genetic susceptibility to SKOA, where a predisposition affects both purposeful and unintentional walking [50]. Neighbourhood environments might also influence physical activity patterns and SKOA risk. Future research could investigate these relationships to better understand complex interactions and inform more effective prevention strategies.

CONCLUSIONS

Increasing daily purposeful walking steps with a higher cadence was associated with a lower risk of SKOA. A daily threshold of 8000 purposeful walking steps emerged as the most optimal for reducing the risk of SKOA in a real-world setting. These findings may promote musculoskeletal health and help

inform future guidelines for knee OA prevention for recommendations based on both step volume and cadence.

Competing interests

All authors have completed the ICMJE uniform disclosure form and declare no financial support from any industry for the submitted work and no other relationships or activities that could appear to have influenced the submitted work.

Contributors

CZ, JW, and GL had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. CZ, JW, and GL are joint corresponding authors. CZ is the lead contact. HH and YW contributed equally to the study and are joint first authors. CZ is the guarantor. Concept and design: JW, CZ, and GL. Acquisition, analysis, or interpretation of data: HH, YW, DH, YZ, WZ, MD, and DW. Drafting of the manuscript: HH, YW, WZ, MD, JW, CZ, and GL. Critical review of the manuscript for important intellectual content: DH, YZ, WZ, MD, DW, JW, CZ, and GL. Statistical analysis: HH and YW. Obtained funding: CZ and GL. Administrative, technical, or material support: JW, CZ, and GL. Supervision: JW, CZ, and GL.

Funding

This work was supported by the National Key Research and Development Program of China (2022YFC3601900 and 2022YFC2505500) and the National Natural Science Foundation of China (81930071, 82072502, U21A20352, 82372474, 82302771). The funding organisations did not have any role in the design and conduct of the study; collection, management, analysis and interpretation of the data; and preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Patient consent for publication

All participants in the UK Biobank study provided written informed consent.

Ethics approval

The UK Biobank study was approved by the National Health Service and the National Research Ethics Service (reference 11/NW/0382).

Provenance and peer review

Not commissioned; externally peer reviewed.

Data availability statement

The statistical code is available from the corresponding authors upon reasonable request. The data may be obtained from the UK Biobank.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.019.

REFERENCES

[1] Cui A, Li H, Wang D, Zhong J, Chen Y, Lu H. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. EClinicalMedicine 2020;29-30:100587. doi: 10.1016/j.eclinm.2020.100587.
[2] Yang G, Wang J, Liu Y, Lu H, He L, Ma C, et al. Burden of knee osteoarthritis in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019. Arthritis Care Res (Hoboken) 2023;75(12):2489–500. doi: 10.1002/acr.25158.
[3] Cherry DK, Hing E, Woodwell DA, Rechtsteiner EA. National Ambulatory Medical Care Survey: 2006 summary. Natl Health Stat Report 2008(3):1–39.
[4] Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019;393 (10182):1745–59. doi: 10.1016/S0140-6736(19)30417-9.
[5] Brisson NM, Gatti AA, Damm P, Duda GN, Maly MR. Association of machine learning-based predictions of medial knee contact force with cartilage loss over 2.5 years in knee osteoarthritis. Arthritis Rheumatol 2021;73(9):1638–45. doi: 10.1002/art.41735.
[6] McPhillips JB, Pellettera KM, Barrett-Connor E, Wingard DL, Criqui MH. Exercise patterns in a population of older adults. Am J Prev Med 1989;5 (2):65–72.
[7] Garduno AC, LaCroix AZ, LaMonte MJ, Dunstan DW, Evenson KR, Wang G, et al. Associations of daily steps and step intensity with incident diabetes in a prospective cohort study of older women: the OPACH study. Diabetes Care 2022;45(2):339–47. doi: 10.2337/dc21-1202.
[8] Saint-Maurice PF, Troiano RP, Jr Bassett DR, Graubard BI, Carlson SA, Shiroma EJ, et al. Association of daily step count and step intensity with mortality among US adults. JAMA 2020;323(12):1151–60. doi: 10.1001/jama.2020.1382.
[9] Lee IM, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019;179(8):1105–12. doi: 10.1001/jamainternmed.2019.0899.
[10] American Heart Association Recommendations for Physical Activity in Adults and Kids 2024 [Internet]. Dallas (TX): American Heart Association; 2024. [cited 2024 Jul 11]. Available from: https://www.heart.org/en/healthy-living/fitness/fitness-basics/aha-recs-for-physical-activity-in-adults.
[11] Hart HF, Gross KD, Crossley KM, Barton CJ, Felson DT, Guermazi A, et al. Step rate and worsening of patellofemoral and tibiofemoral joint osteoarthritis in women and men: the Multicenter Osteoarthritis Study. Arthritis Care Res (Hoboken) 2020;72(1):107–13. doi: 10.1002/acr.23864.
[12] Master H, Thoma LM, Neogi T, Dunlop DD, LaValley M, Christiansen MB, et al. Daily walking and the risk of knee replacement over 5 years among adults with advanced knee osteoarthritis in the United States. Arch Phys Med Rehabil 2021;102(10):1888–94. doi: 10.1016/j.apmr.2021.05.014.
[13] Zhu Z, Aitken D, Cicuttini F, Jones G, Ding C. Ambulatory activity interacts with common risk factors for osteoarthritis to modify increases in MRI-detected osteophytes. Osteoarthritis Cartilage 2019;27(4):650–8. doi: 10.1016/j.joca.2018.12.023.
[14] Lo GH, Vinod S, Richard MJ, Harkey MS, McAlindon TE, Kriska AM, et al. Association between walking for exercise and symptomatic and structural progression in individuals with knee osteoarthritis: data from the osteoarthritis initiative cohort. Arthritis Rheumatol 2022;74(10):1660–7. doi: 10.1002/art.42241.
[15] Lin W, Alizai H, Joseph GB, Srikhum W, Nevitt MC, Lynch JA, et al. Physical activity in relation to knee cartilage T2 progression measured with 3 T MRI over a period of 4 years: data from the Osteoarthritis Initiative. Osteoarthritis Cartilage 2013;21(10):1558–66. doi: 10.1016/j.joca.2013.06.022.
[16] Brisson NM, Gatti AA, Maly MR. Association of pain and steps per day in persons with mild-to-moderate, symptomatic knee osteoarthritis: a mixed-effects models analysis of multiple measurements over three years. Arthritis Care Res (Hoboken) 2020;72(1):114–21. doi: 10.1002/acr.23842.
[17] Øiestad BE, Quinn E, White D, Roemer F, Guermazi A, Nevitt M, et al. No association between daily walking and knee structural changes in people at risk of or with mild knee osteoarthritis. Prospective data from the Multicenter Osteoarthritis Study. J Rheumatol 2015;42(9):1685–93. doi: 10.3899/jrheum.150071.
[18] Gates LS, Perry TA, Golightly YM, Nelson AE, Callahan LF, Felson D, et al. Recreational physical activity and risk of incident knee osteoarthritis: an international meta-analysis of individual participant-level data. Arthritis Rheumatol 2022;74(4):612–22. doi: 10.1002/art.42001.
[19] Brisson NM, Wiebenga EG, Stratford PW, Beattie KA, Totterman S, Tamez-Peña JG, et al. Baseline knee adduction moment interacts with body mass index to predict loss of medial tibial cartilage volume over 2.5 years in knee osteoarthritis. J Orthop Res 2017;35(11):2476–83. doi: 10.1002/jor.23564.
[20] Wang Y, Simpson JA, Wluka AE, Teichtahl AJ, English DR, Giles GG, et al. Is physical activity a risk factor for primary knee or hip replacement due to osteoarthritis? A prospective cohort study. J Rheumatol 2011;38(2):350–7. doi: 10.3899/jrheum.091138.

[21] Doré DA, Winzenberg TM, Ding C, Otahal P, Pelletier JP, Martel-Pelletier J, et al. The association between objectively measured physical activity and knee structural change using MRI. Ann Rheum Dis 2013;72(7):1170–5. doi: 10.1136/annrheumdis-2012-201691.

[22] Tudor-Locke C, Han H, Aguiar EJ, Barreira TV, Schuna Jr JM, Kang M, et al. How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. Br J Sports Med 2018;52(12):776–88. doi: 10.1136/bjsports-2017-097628.

[23] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203–9. doi: 10.1038/s41586-018-0579-z.

[24] Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc 2005;37(11 Suppl):S531–43. doi: 10.1249/01.mss.0000185657.86065.98.

[25] Rowlands AV, Maylor B, Dawkins NP, Dempsey PC, Edwardson CL, Soczawa-Stronczyk AA, et al. Stepping up with GGIR: validity of step cadence derived from wrist-worn research-grade accelerometers using the verisense step count algorithm. J Sports Sci 2022;40(19):2182–90. doi: 10.1080/02640414.2022.2147134.

[26] Funck-Brentano T, Nethander M, Movérare-Skrtic S, Richette P, Ohlsson C. Causal factors for knee, hip, and hand osteoarthritis: a Mendelian randomization study in the UK Biobank. Arthritis Rheumatol 2019;71(10):1634–41. doi: 10.1002/art.40928.

[27] First Occurrence of Health Outcomes Defined by 3-character ICD10 code. UK Biobank; 2019 [cited 2023 Oct 23]. Available from: https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/first_occurrences_outcomes.pdf.

[28] Hanlon P, Nicholl BI, Jani BD, Lee D, McQueenie R, Mair FS. Frailty and pre-frailty in middle-aged and older adults and its association with multimorbidity and mortality: a prospective analysis of 493 737 UK Biobank participants. Lancet Public Health 2018;3(7):e323–32. doi: 10.1016/S2468-2667(18)30091-4.

[29] Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79. doi: 10.1186/1479-5868-8-79.

[30] Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009;28(25):3083–107. doi: 10.1002/sim.3697.

[31] Brittain EL, Han L, Annis J, Master H, Hughes A, Roden DM, et al. Physical activity and incident obesity across the spectrum of genetic risk for obesity. JAMA Netw Open 2024;7(3):e243821. doi: 10.1001/jamanetworkopen.2024.3821.

[32] Williams PT. Reduced diabetic, hypertensive, and cholesterol medication use with walking. Med Sci Sports Exerc 2008;40(3):433–43. doi: 10.1249/MSS.0b013e31815f38f1.

[33] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94(446):496–509. doi: 10.2307/2670170.

[34] Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. Stat Med 2010;29(9):1037–57. doi: 10.1002/sim.3841.

[35] VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017;167(4):268–74. doi: 10.7326/M16-2607.

[36] Fenton SAM, Neogi T, Dunlop D, Nevitt M, Doherty M, Duda JL, et al. Does the intensity of daily walking matter for protecting against the development of a slow gait speed in people with or at high risk of knee osteoarthritis? An observational study. Osteoarthritis Cartilage 2018;26(9):1181–9. doi: 10.1016/j.joca.2018.04.015.

[37] White DK, Tudor-Locke C, Zhang Y, Fielding R, LaValley M, Felson DT, et al. Daily walking and the risk of incident functional limitation in knee osteoarthritis: an observational study. Arthritis Care Res (Hoboken) 2014;66(9):1328–36. doi: 10.1002/acr.22362.

[38] Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007;298(19):2296–304. doi: 10.1001/jama.298.19.2296.

[39] Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman B, Aliabadi P, et al. Risk factors for incident radiographic knee osteoarthritis in the elderly: the Framingham Study. Arthritis Rheum 1997;40(4):728–33. doi: 10.1002/art.1780400420.

[40] Bassett DR, Schneider PL, Huntington GE. Physical activity in an Old Order Amish community. Med Sci Sports Exerc 2004;36(1):79–85. doi: 10.1249/01.MSS.0000106184.71258.32.

[41] Mow VC, Ratcliffe A, Poole AR. Cartilage and diarthrodial joints as paradigms for hierarchical materials and structures. Biomaterials 1992;13(2):67–97. doi: 10.1016/0142-9612(92)90001-5.

[42] Voinier S, Moore AC, Benson JM, Price C, Burris DL. The modes and competing rates of cartilage fluid loss and recovery. Acta Biomater 2022;138:390–7. doi: 10.1016/j.actbio.2021.11.014.

[43] Moore AC, Burris DL. Tribological rehydration of cartilage and its potential role in preserving joint health. Osteoarthritis Cartilage 2017;25(1):99–107. doi: 10.1016/j.joca.2016.09.018.

[44] Huang AH, Farrell MJ, Kim M, Mauck RL. Long-term dynamic loading improves the mechanical properties of chondrogenic mesenchymal stem cell-laden hydrogel. Eur Cell Mater 2010;19:72–85. doi: 10.22203/ecm.v019a08.

[45] Soltz MA, Ateshian GA. Interstitial fluid pressurization during confined compression cyclical loading of articular cartilage. Ann Biomed Eng 2000;28(2):150–9. doi: 10.1114/1.239.

[46] Linn FC. Lubrication of animal joints. I. The arthrotripsometer. J Bone Joint Surg Am 1967;49(6):1079–98.

[47] Kettle VE, Madigan CD, Coombe A, Graham H, Thomas JJC, Chalkley AE, et al. Effectiveness of physical activity interventions delivered or prompted by health professionals in primary care settings: systematic review and meta-analysis of randomised controlled trials. BMJ 2022;376:e068465. doi: 10.1136/bmj-2021-068465.

[48] Quicke JG, Foster NE, Thomas MJ, Holden MA. Is long-term physical activity safe for older adults with knee pain?: a systematic review. Osteoarthritis Cartilage 2015;23(9):1445–56. doi: 10.1016/j.joca.2015.05.002.

[49] Kolasinski SL, Neogi T, Hochberg MC, Oatis C, Guyatt G, Block J, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020;72(2):220–33. doi: 10.1002/art.41142.

[50] Boer CG, Hatzikotoulas K, Southam L, Stefánsdóttir L, Zhang Y, Coutinho de Almeida R, et al. Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell 2021;184(24):6003–5. doi: 10.1016/j.cell.2021.11.003.

Ann Rheum Dis 84 (2025) 643–652

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Original research

Systematic literature review and meta-analysis informing the EULAR points to consider on the initiation of targeted therapies in patients with inflammatory arthritis and a history of cancer

Eden Sebbag[1], Juan Molina-Collada[2], Ramatoulaye Ndoye[3], Daniel Aletaha[4], Johan Askling[5], Karolina Gente[6], Heidi Bertheussen[7], Samuel Bitoun[8], Ertugrul Cagri Bolek[9], Maya H. Buch[10], Gerd R. Burmester[11], Helena M. Canhão[12], Katerina Chatzidionysiou[13], Jeffrey R. Curtis[14], Francois-Xavier Danlos[15,16], Vera Guimarães[17], Merete Lund Hetland[18,19], Florenzo Iannone[20], Marie Kostine[21], Tue Wenzel Kragstrup[22,23], Tore K. Kvien[24], Anne Constanze Regierer[25], Hendrik Schulze-Koops[26], Nathanaël Sedmak[3], Lucía Silva-Fernández[27], Zoltan Szekanecz[28], Kim Lauper[29], Axel Finckh[29], Jacques-Eric Gottenberg[1,*]

[1] *Department of Rheumatology, Centre National de Référence des Maladies Auto-Immunes Rares, Hôpitaux Universitaires de Strasbourg, Université de Strasbourg, Strasbourg, France*
[2] *Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Instituto de Investigación Sanitaria Gregorio Marañón (IiSGM), Madrid, Spain*
[3] *Département de Biostatistique, Les Hopitaux Universitaires de Strasbourg Hopital de Hautepierre, Strasbourg, France*
[4] *Department of Rheumatology, Medical University of Vienna, Vienna, Austria*
[5] *Karolinska Institutet, Stockholm, Sweden*
[6] *Department of Hematology, Oncology and Rheumatology, University Hospital Heidelberg, Heidelberg, Baden-Württemberg, Germany*
[7] *Patient research partner, Oslo, Norway*
[8] *Department of Rheumatology, Université Paris-Saclay, Assistance Publique Hôpitaux de Paris, Hôpital Bicêtre, INSERM, CEA, UMR 1184, FHU CARE, Le Kremlin Bicêtre, France*
[9] *Division of Rheumatology, Department of Internal Medicine, Gazi University Faculty of Medicine, Ankara, Turkey*
[10] *Centre for Musculoskeletal Research, The University of Manchester, Manchester, UK*
[11] *Department of Rheumatology and Clinical Immunology, Charité - Universitätsmedizin, Berlin, Germany*
[12] *CHRC, NOVA Medical School, Lisboa, Portugal*
[13] *Department of Medicine, Karolinska Institutet, Stockholm, Sweden*
[14] *Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, Alabama, USA*
[15] *Drug Development Department, INSERM U1015, Gustave Roussy, Villejuif, France*
[16] *Department of Medical Oncology, Dana Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA*
[17] *Liga Portuguesa Contra as Doenças Reumáticas, Lisbon, Portugal*

*Correspondence to Dr. Jacques-Eric Gottenberg.
E-mail address: jacques-eric.gottenberg@chru-strasbourg.fr (J.-E. Gottenberg).
Social media: @jmolinacollada (J. Molina-Collada), @ertugrul_cagri (E. C. Bolek), @MarieKostine (M. Kostine), @KragstrupTW (T. W. Kragstrup), @k_lauper (K. Lauper)

Handling editor Kimme L. Hyrich.

https://doi.org/10.1136/ard-2024-225981

[18] Copenhagen Center for Arthritis Research (COPECARE), Center for Rheumatology and Spine Diseases, Centre of Head and Orthopedics, Copenhagen University Hospital Rigshospitalet, Glostrup, Denmark
[19] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[20] Department of Rheumatology, University of Bari Aldo Moro, Bari, Italy
[21] Department of Rheumatology, Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France
[22] Department of Biomedicine, Aarhus University, Aarhus C, Denmark
[23] Department of Rheumatology, Silkeborg Regional Hospital and Aarhus University Hospital, Aarhus, Denmark
[24] Center for Treatment of Rheumatic and Musculoskeletal Diseases (REMEDY), Diakonhjemmet Hospital, Oslo, Norway
[25] Epidemiology and Health Care Research, German Rheumatism Research Center (DRFZ Berlin), Berlin, Germany
[26] Division of Rheumatology and Clinical Immunology, Internal Medicine IV, LMU Klinikum, Munchen, Germany
[27] Department of Rheumatology, Complexo Hospitalario Universitario de A Coruña, A Coruña, Spain
[28] Department of Rheumatology, University of Debrecen, Debrecen, Hungary
[29] Division of Rheumatology & Geneva Centre for Inflammation research/ Department of Medicine, Geneva University Hospitals, Faculty of Medicine, University of Geneva, Genève, Switzerland

ARTICLE INFO

Article history:
Received 16 April 2024
Accepted 5 November 2024

Keywords:
Arthritis
Autoimmunity
Inflammation
Biological Therapy

ABSTRACT

Background: Targeted therapies have been associated with potential risk of malignancy, which is a common concern in daily rheumatology practice in patients with inflammatory arthritis (IA) and a history of cancer.
Objectives: To perform a systematic literature review to inform a Task Force formulating EULAR points to consider on the initiation of targeted therapies in patients with IA and a history of cancer.
Methods: Specific research questions were defined within the Task Force before formulating the exact research queries with a librarian. We included studies reporting a relative risk measure of patients with a history of cancer initiating a targeted therapy or a conventional synthetic disease-modifying antirheumatic drug (csDMARD), regardless of the time since diagnosis of cancer. All relevant studies included in PubMed or Embase up to 15 July 2022 were included. Two reviewers independently performed standardised article selection, data extraction, synthesis and risk of bias assessment.
Results: 14 published articles and one ACR abstract fulfilled the inclusion criteria. All studies were high-quality observational studies, representing a median follow-up from treatment initiation of 4.52 years among 4428 patients and 15 062 patient-years of follow-up for new or recurrent cancer.
All patients had a history of cancer, most frequently solid cancer, most frequently receiving treatment for rheumatoid arthritis and most frequently treated with tumour necrosis factor-alpha inhibitors. Across these studies, the overall HR of new incident cancer or cancer recurrence was 0.90 (95% CI 0.74 to 1.10) for patients receiving a targeted therapy versus a csDMARD.
Conclusion: Overall, the targeted therapies and clinical contexts covered by the included studies were not associated with an increased risk of new incident cancer or cancer recurrence as compared with csDMARDs.

INTRODUCTION

Targeted therapies, including biologic and targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs), have considerably improved the long-term outcomes of patients with inflammatory arthritis (IA), improving quality of life and with remission as an attainable goal [1–3]. Potential associations between targeted therapies in patients with IA and malignancy are a frequent concern in daily rheumatology practice. Cancer risks may be increased in patients with rheumatoid arthritis (RA) and other chronic inflammatory diseases as compared with the general population [4,5]. Because immunity plays an important role in tumour immunosurveillance, the use of b/tsDMARDs is a concern, also in the context of treating RA in individuals with a history of cancer [6]. Improvements in cancer therapy have led to major gains in survival.

For all the above reasons, including an ageing population, the number of patients with a history of cancer and a need for treatment with b/tsDMARDs is increasing. At the same time, the evidence to guide b/tsDMARD treatment in this context is limited. Patients with a history of cancer are routinely excluded from randomised controlled trials evaluating targeted therapies, and there are limited data on the risk of a new cancer or cancer recurrence in this setting. A recent study suggested that the reluctance to use a targeted therapy in patients with RA and a history of cancer might result in undertreatment of some patients and increased use of rituximab due to channelling bias [7].

To inform the Task Force responsible for the EULAR points to consider (PTC) on the initiation of targeted therapies in patients with IA and a history of cancer, we performed a systematic literature review (SLR) and meta-analysis to investigate the risk of cancer in patients with IA treated with targeted therapies and a history of cancer.

METHODS

Literature search

At the first meeting, the Task Force for the EULAR PTC on the initiation of targeted therapies in patients with IA and a history

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Potential associations between targeted therapies in patients with inflammatory arthritis (IA) and malignancy are a frequent concern in daily rheumatology practice.
- No specific framework has been proposed to weigh the benefit/risk balance of initiating or reinitiating a targeted therapy in patients with IA and a history of cancer.

WHAT THIS STUDY ADDS

- This systematic literature review (SLR) and meta-analysis did not demonstrate any increased risk of new incident cancer or cancer recurrence in patients with IA who were treated with targeted therapy compared with those who were treated differently after a history of cancer.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- The results of this SLR can help clinicians improve therapeutic decision-making in the context of a patient with a history of cancer-initiating a targeted therapy for IA. The results have informed an international Task Force formulating EULAR points to consider.

of cancer defined the research points under supervision of one methodologist (AF) and a co-methodologist (KL). It also outlined the scope of the literature search according to the pre-specific Population, Intervention, Comparator and Outcomes (PICOs) format questions and defined the criteria for an eligible study. In a first meeting, the EULAR Task Force agreed to focus the SLR on clinical data in patients receiving any targeted therapy for an inflammatory or autoimmune rheumatic or skin or bowel disease, with a history of cancer regardless of the time since diagnosis of cancer.

The Task Force decided to include dermatological and gastrointestinal immune-mediated diseases, as these conditions are treated with the same therapies. The search was performed in PubMed via MEDLINE and Embase without language restrictions from 1 January 2010 to 15 July 2022 and included recent abstracts from large international congresses. Details on complete search strategies are in figure 1. The detailed search strategy is presented in online supplemental table 1.

Inclusion criteria for studies included the reporting of an outcome measure of the risk of cancer in patients with a history of cancer who initiated a targeted therapy for an inflammatory rheumatic, bowel or skin disease, regardless of the time since diagnosis of cancers. For the main analysis, studies were eligible if they included a comparator group and reported a relative risk measure (eg, HR) of new cancer or cancer recurrence between groups. Studies reporting incidence data without HRs were not included in the meta-analysis. In the case of several publications from the same registry, only the most recent article was selected, unless the main outcome was specified to be different (eg, new breast cancer/breast cancer recurrence in patients with a history of breast cancer, new non-melanoma skin cancer (NMSC)/NMSC recurrence in patients with a history of NMSC). Stratified analyses considered the histological type of the initial cancer, the time from the initial cancer diagnosis to the initiation of targeted therapy, and the duration of targeted therapy. Several sensitivity analyses were performed to study specific treatments, to analyse only the most recent study from registries that had published multiple reports, and to analyse only patients with RA.

Selection of studies, data extraction and assessment of risk of bias

Two independent reviewers (ES and JM-C) screened titles and abstracts to assess eligibility. Subsequently, all potentially eligible articles were read in full text to decide whether they fulfilled the inclusion criteria. The reviewers used a standardised data extraction form with the Rayyan online tool [8] to extract data from eligible studies regarding study and population

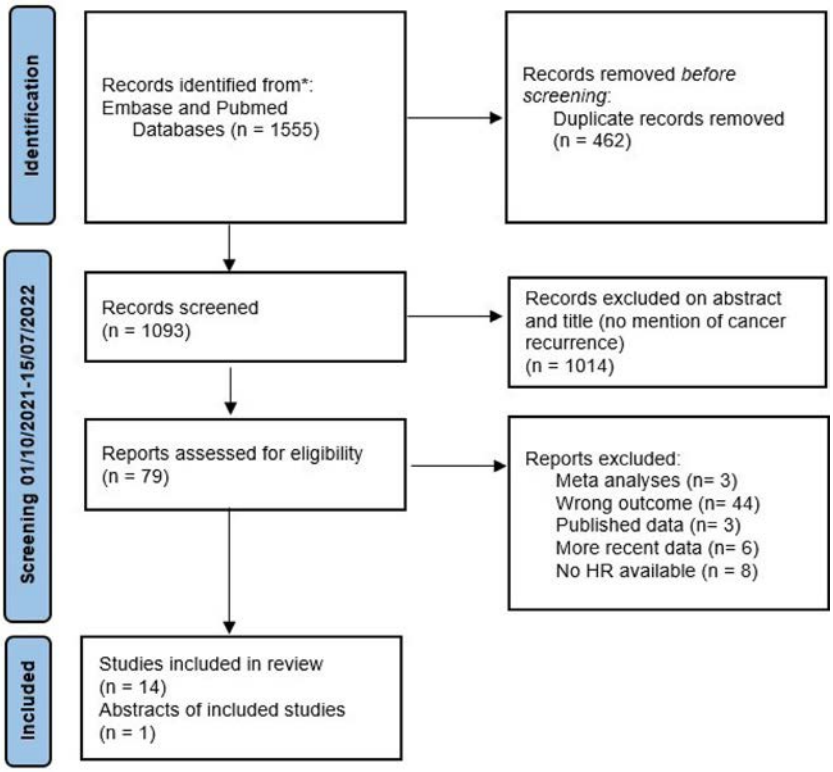


Figure 1. Flow chart of the study.

characteristics, inclusion/exclusion criteria, drug exposure time, interventions and outcome definition. Any disagreement between reviewers was resolved by discussion. Tables including the summary of findings of the studies included were generated. The risk of bias (RoB) of each included study was assessed with the ROBINS-I tool [9]. The quality of the studies was graded according to the Newcastle-Ottawa Scale for quality assessment. Discrepancies between reviewers regarding study selection, data extraction and RoB assessment were resolved by discussion with the methodologists.

Statistical analysis

The included studies were used for a meta-analysis to assess the association between targeted therapy (b/tsDMARD) and new cancer or cancer recurrence after prior cancer, the primary outcome of all included studies. Heterogeneity between studies was assessed with the I^2 calculation. Heterogeneity as well as differences in the treatment used for every study (tumour necrosis factor (TNF) inhibitor, rituximab, vedolizumab or other) led to the choice of a random-effects model to produce a combined relative risk and its corresponding 95% CI. Given the limited number of studies available for analysis, the Hartung-Knapp-Sidik-Jonkman approximation was employed as a means of mitigating potential bias. Publication bias was assessed with a funnel plot. When needed, the statistical results of the studies were pooled and weighted according to their sample size. All analyses were performed with R V.4.1.2. $P<0.05$ was considered statistically significant.

RESULTS

Primary analysis

Fifteen articles were included in the primary analysis, representing (disregarding any overlap between studies) 4428 patients and 15 062 patient-years in the group receiving targeted therapy, and 13 698 patients and 41 160 patient-years in the control group (typically receiving a csDMARD; table 1). The flow chart of articles in the SLR is available in figure 1. Nine studies used data from European registries, four studies used data from US healthcare databases, one study was a retrospective single-centre study and one study was a retrospective multicentric study. According to the Newcastle-Ottawa Scale, the overall quality of the studies was considered good, with a calculated score of >6 for all studies (online supplemental table 1).

The median age at cancer diagnosis was 52.5 years (min −max range 49−57). In the 13 studies reporting the follow-up duration, the median follow-up was 4.52 years (min−max range 9.8 months−8.8 years). The reported median time from the index cancer to the initiation of b/tsDMARD treatment was 4 years. Six out of fifteen studies adjusted the reported HR for general risk factors such as smoking, age and sex, and four studies additionally adjusted for specific cancer risk factors such as cancer type and index cancer stage.

For all included studies, the primary objective was to analyse cancer recurrence or new incident cancer in patients with a history of cancer-initiating a targeted therapy. In most studies (>90%), previous cancers were solid tumours. Two studies focused on breast cancer [10,11], one study on melanoma [12], two studies on NMSC [13,14] and one study on squamous cell carcinomas of the head and neck [15].

12 studies included patients with RA exclusively and 2 studies included patients with inflammatory bowel disease (IBD). One study included RA, IBD and psoriasis. The details of the targeted therapies included are in table 2.

Most of the b/tsDMARDs studied evaluated TNF inhibitors (4162 patients and 13 519 patient-years), but data on other bDMARDs were also studied: vedolizumab (130 patients and 1087 patient-years), ustekinumab (66 patients and 228 patient-years) and rituximab (100 patients and 261 patient-years). No studies included tsDMARDs.

In total, 460 new cancers or cancer recurrences (46.2/1000 patient-years) were reported in the bDMARD-treated group: 428 in the TNF-inhibitor group (47.6/1000 patient-years), 9 in the rituximab group (35.7/1000 patient-years), 19 in the vedolizumab group (17.1/1000 patient-years) and 3 in the ustekinumab group. A total of 1394 new cancers or cancer recurrences were reported in the control csDMARD group (42.3/1000 patient-years, 423/100 patient-years). The pooled HR for new incident cancer or cancer recurrence between the bDMARD and control group was 0.90 (95% CI 0.74 to 1.10) (figure 2).

In analyses by drug exposure, the pooled adjusted HR was 0.94 (95% CI 0.76 to 1.18)) for the TNF inhibitor group, 0.81 (0.35 to 1.89) for the rituximab group (figure 3) and 0.49 (0.14 to 1.65) for the vedolizumab group (online supplemental figure 1).

Stratified analyses

Stratified analyses considered the histological type of initial cancer, the time from the initial cancer diagnosis to the initiation of targeted therapy and the duration of targeted therapy.

Six studies specifically studied patients with a history of solid cancer (excluding NMSCs and melanomas). All compared the occurrence of new incident cancer or cancer recurrence in patients with a history of solid cancer who received a TNF inhibitor or csDMARD. Two studies specifically studied breast cancer [10,11], one study focused on squamous cell carcinomas of the head and neck [15] and three studies evaluated different types of solid cancers [16−18]. The pooled adjusted random effects meta-analysis HR for new cancer/cancer recurrence was 0.89 (95% CI 0.66 to 1.21) for patients with a history of solid cancer who received a TNF inhibitor versus a csDMARD (online supplemental figure 1).

Three studies specifically analysed patients with a history of NMSC [13,14,16] who received a TNF inhibitor. The pooled adjusted random effects meta-analysis HR was 1.23 (95% CI 0.90 to 1.70) as compared with csDMARD users (online supplemental figure 3).

Five studies [17,19−22], reported a median time from the initial cancer diagnosis to the initiation of bDMARDs of less than 5 years. One study focused on patients with an IBD, and the other four studies focused on patients with RA. One study [21] reported data with vedolizumab and TNF inhibitors, the other studies reported on TNF inhibitors only. When analysing the risk of new incident cancer or cancer recurrence on bDMARDs initiated within 5 years of cancer diagnosis, the pooled adjusted random effects meta-analysis HR was 0.88 (95% CI 0.64 to 1.21) as compared with csDMARD users (figure 4). Three studies [11,17,23] reported a time from the initial cancer diagnosis to the initiation of biological therapy of more than 5 years. One study [23] reported rituximab and TNF-inhibitor data; the other two studies reported TNF-inhibitor data only. When analysing the risk of cancer recurrence on bDMARDs initiated after more than 5 years of cancer diagnosis, the pooled adjusted random effects meta-analysis HR was 0.82 (95% CI 0.47 to 1.44) as compared with csDMARD users (online supplemental figure 4).

Table 1

Details of the studies included in the systematic literature review

Author	Year	Country	Group	Number	Patient-years	Cancer endpoint	Inflammatory disease studied	Events*	Rate/1000 patient-years	HR (95% CI)	Crude HR	Adjusted HR	Adjustments
Waljee *et al*16	2019	Denmark	TNF inhibitors	434	2376	Mixed cancer type	RA, Pso, IBD	72	30.3	0.82 (0.51 to 1.11)	0.86 (0.66 to 1.12)	0.82 (0.51 to 1.11)	(1)
			Biologic-naive	4328	16 376			563	34.4	Reference			
Mamtani *et al* [10]	2016	USA	TNF inhibitors	291	764	Breast cancer	RA	17	22.3	1.13 (0.65 to 1.97)		1.13 (0.65 to 1.97)	No covariates needed
			csDMARDs	1164	2466			48	19.5	Reference			
Strangfeld *et al* [19]	2010	Germany	TNF inhibitors	59	198	Mixed cancer type	RA	9	45.5	1.4 (0.5 to 5.55)	1.4 (0.5 to 5.55)		None
			csDMARDs	56	159			5	31.4	Reference			
Strangfeld *et al* [22]	2013	Germany	Rituximab	77	180	Mixed cancer type	RA	7	38.9	1.1 (0.4 to 2.7)	1.1 (0.4 to 2.7)		None
			csDMARDs	112	361			13	36	Reference			
Scott *et al*[13]	2015	USA	TNF inhibitors	1839	1465	NMSC	RA	109	74.4	1.49 (1.03 to 2.16)		1.49 (1.03 to 2.16)	(2)
			csDMARDs	4414	4631			335	72.3	Reference			
Silva-Fernández *et al*[23]	2016	UK	TNF inhibitors	243	1591	Mixed cancer type	RA	53	33.3	0.56 (0.36 to 0.88)	0.51 (0.33 to 0.79)	0.56 (0.36 to 0.88)	(3)
			Rituximab	23	81			2	24.7	0.44 (0.11 to 1.82)	0.45 (0.11 to 1.97)	0.44 (0.11 to 1.82)	
			csDMARDs	159	855			46	53.8	Reference			
Raaschou *et al* [17]	2018	Sweden	TNF inhibitors	467	2471	Mixed cancer type	RA	42	17.0	1.06 (0.73 to 1.54)		1.06 (0.73 to 1.54)	(4)
			Biologic-naive	2164	9394			155	16.5	Reference			
Philipps *et al* [15]	2015	USA	TNF inhibitors	40	256	Head and neck cancer	RA	7	27.3	0.9 (0.4 to 2.1)	0.9 (0.4 to 2.1)		None
			csDMARDs	190	1195			35	29.3	Reference			
Aaltonen *et al* [18]	2015	Finland	TNF inhibitors	100	233	Mixed cancer type	RA	3	12.9	2.2 (0.2 to 20.7)	2.2 (0.2 to 20.7)		None
			csDMARDs	77	169			1	5.9	Reference			
Raaschou *et al* [12]	2013	Sweden	TNF inhibitors	54	271	Melanoma	RA	3	11.1	3.2 (0.8 to 13.1)		3.2 (0.8 to 13.1)	(5)
			Biologic-naive	295	1370			10	7.3	Reference			
Mercer *et al* [14]	2012	UK	TNF inhibitors	177	627	NMSC	RA	29	46.3	0.7 (0.26 to 1.94)		0.7 (0.26 to 1.94)	(6)
			Biologic-naive	106	276			23	79.9	Reference			
Raaschou *et al* [11]	2014	Sweden	TNF inhibitors	120	592	Breast cancer	RA	9	15	1.1 (0.4 to 2.8)	0.8 (0.3 to 2.1)	1.1 (0.4 to 2.8)	(7)
			csDMARDs	120	550			9	16	Reference			
Axelrad *et al* [20]	2016	USA	TNF inhibitors	55	285	Mixed cancer type	IBD	7	24.6	0.35 (0.09 to 1.09)		0.35 (0.09 to 1.09)	(8)
			csDMARDs	149	852			46	852	Reference			
Vedamurthy *et al*[21]	2022	USA	Vedolizumab	96	821	Mixed cancer type	IBD	18	22	0.72 (0.38 to 1.39)		0.72 (0.38 to 1.39)	(9)
			TNF inhibitors	184	1452			61	42	1.03 (0.65 to 1.64)		1.03 (0.65 to 1.64)	
			csDMARDs	183	1378			78	56	Reference			
Hasan *et al*[34]	2022	Canada	Vedolizumab	34	266	Mixed cancer type	IBD	1	4	0.18 (0.03 to 1.35)		0.18 (0.03 to 1.35)	(10)
			Ustekinumab	27	164			3	18	0.88 (0.25 to 3.03)		0.88 (0.25 to 3.03)	
			TNF inhibitors	99	938			7	7	0.47 (0.20 to 1.12)		0.47 (0.20 to 1.12)	

Adjustments when done: (1) adjusted for sex; initial primary cancer type; diagnosis of inflammatory bowel disease (IBD), rheumatoid arthritis or psoriasis; age at diagnosis of IBD, rheumatoid arthritis or psoriasis; age at initial cancer diagnosis; number of baseline hospital admissions; and number of baseline outpatient visits; (2) TNF exposure; (3) age-, sex- and smoking status-adjusted HR; (4) adjusted for the matching variables sex, birth year (±10 years), year of diagnosis (±5 years) of the index cancer, cancer type and index cancer stage; (5) age and sex; (6) age, sex, smoking, exposure to cyclosporine or azathioprine, NSAIDs and baseline characteristics; (7) adjusted for breast cancer characteristics (nodal state, type of surgery, chemotherapy) and comorbidities (diabetes mellitus, ischaemic heart disease, chronic obstructive pulmonary disease and joint surgery); (8) adjusted by recurrence risk type of prior cancer; (9) adjusted for age at index cancer, IBD subtype, smoking history, antimetabolite exposure, cancer category, cancer stage and time to biological. Time to biological for the no immunosuppressant group was set as the median of the anti-TNF group; (10) adjusted for recurrence risk, type of primary cancer, baseline demographics and disease characteristics.

* New incident cancer or recurrence of a previous cancer.csDMARD, conventional synthetic disease-modifying antirheumatic drug; IBD, inflammatory bowel disease; NMSC, non-melanoma skin cancer; Pso, psoriasis; RA, rheumatoid arthritis; TNF, tumour necrosis factor.

Table 2
Details of the targeted therapies included in the study

Targeted therapy	Patients	Patient-years
TNF inhibitors	4162	13 519
Vedolizumab	130	1087
Ustekinumab	66	228
Rituximab	100	261

TNF, tumour necrosis factor.

Four studies [12,15,17,23] reported a mean duration of bDMARD exposure exceeding 5 years. All these studies focused on treatment with TNF inhibitors. The pooled adjusted random effects meta-analysis HR of new incident cancer or cancer recurrence was 0.87 (95% CI 0.63 to 1.19) as compared with csDMARD users (online supplemental figure 5). For the 11 studies with a duration of TNF-inhibitor exposure of less than 5 years, the pooled adjusted random effects meta-analysis HR was 0.93 (95% CI 0.71 to 1.20) as compared with csDMARD users (online supplemental figure 6).

Sensitivity analyses and RoB assessment

Furthermore, several sensitivity analyses were conducted.

A sensitivity analysis was conducted on the data set comprising only patients with IA (12 studies), all of them having RA. The overall risk of new incident cancer or cancer recurrence in RA patients with a history of cancer treated with bDMARDs versus csDMARDs was 1.03 (95% CI 0.79 to 1.34) (figure 5).

In another sensitivity analysis, we retained only the most recent study from each registry to limit the risk of taking into account the same patients several times when the same registry had published multiple reports. The following studies were

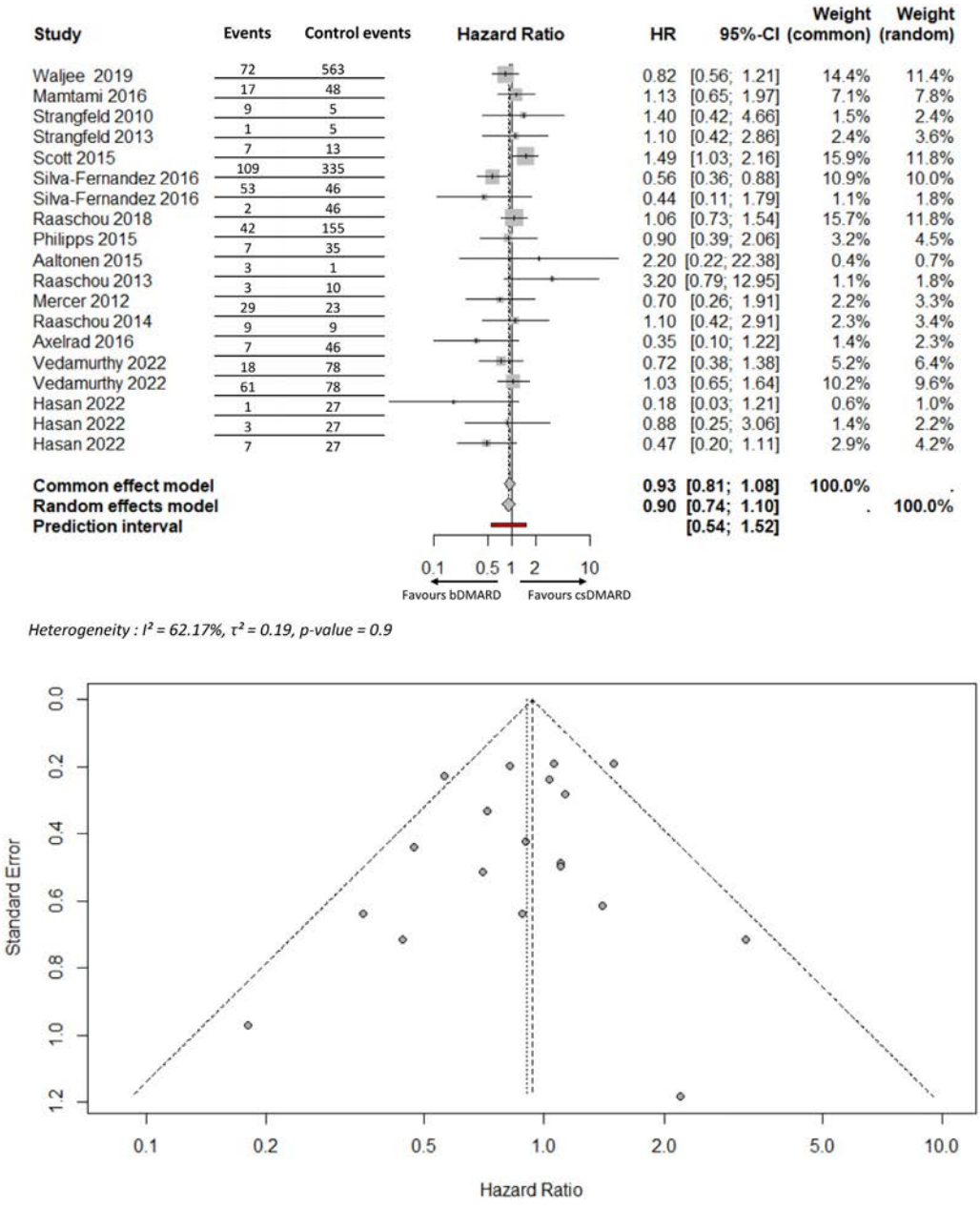


Figure 2. Overall risk of new cancer or cancer recurrence for patients receiving biologic disease-modifying anti-rheumatic drugs (bDMARDs) versus conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) with the respective funnel plot of the studies. All bDMARDs versus csDMARDs in patients with a history of cancer.

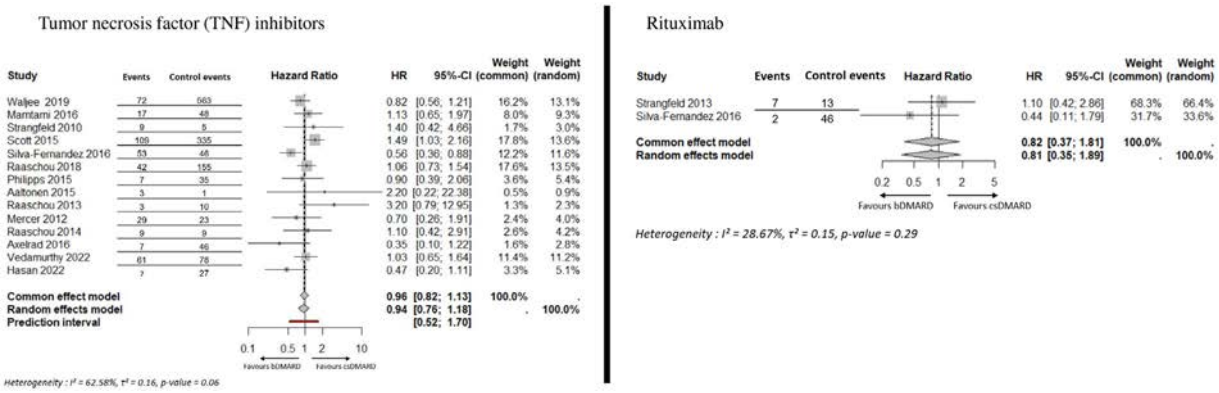


Figure 3. Overall risk of new cancer or cancer recurrence in patients receiving biologic disease-modifying antirheumatic drugs (bDMARDs) versus conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), stratified by type of bDMARD used (TNF inhibitors or rituximab).

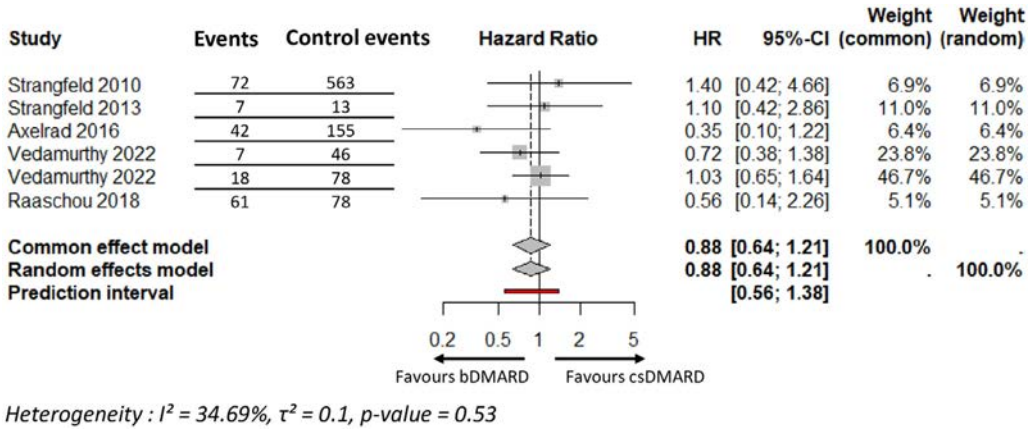


Figure 4. Risk of new cancer or cancer recurrence for patients receiving biologic disease-modifying antirheumatic drugs (bDMARDs) versus conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) less than 5 years after initial cancer diagnosis. Use of bDMARD within 5 years of cancer diagnosis.

removed in this sensitivity analysis: Strangfeld *et al* [22] and Raaschou *et al* [11,12]. The overall risk of new incident or recurrent cancer in the sensitivity analysis was 0.87 (95% CI 0.70 to 1.09) (online supplemental figure 7).

In the last sensitivity analysis, we combined only RA patients with the latest registry data available. The pooled adjusted HR of new incident cancer or cancer recurrence in this sensitivity analysis was 0.99 (95% CI 0.72 to 1.37) (online supplemental figure 8).

The RoB, evaluated with the ROBINS-I tool, was considered low. However, heterogeneity could have been impacted by a lot more factors (besides the ones included in the I^2 calculation): the studies were intrinsically heterogeneous because they addressed different treatments and different diseases, and some of them included a limited sample size, which affects the I^2 calculation.

DISCUSSION

In this SLR and meta-analysis including a total of 15 studies covering 4428 patients and 15 062 patient-years in the groups receiving targeted therapies and 13 698 patients and 41 160 patient-years in the control groups, we found no indication of increased risk of new cancer or cancer recurrence in patients receiving targeted therapy.

The SLR aimed to analyse new incident cancer or cancer recurrence (relapses of a previous cancer) not only in IA but also in inflammatory skin and bowel diseases, yet it mainly identified data for patients with IA. Among patients with IA, all patients had RA. The systematic review aimed to collect data on all targeted therapies, but most studies enrolled only patients with a history of cancer receiving TNF inhibitors. No study analysed patients for whom the cancer was not in remission or presumed to be in remission at the time of the initiation of targeted therapy. This study aimed to assess the incidence of cancer between targeted therapies and csDMARDs. However, some csDMARDs might increase the risk of cancer in patients as compared with healthy individuals, such as methotrexate for NMSCs [24]. Independent of treatments, some rheumatic diseases are associated with increased risk of some cancers [25,26], notably via disease activity, autoimmunity, inflammation and comorbidities such as smoking. Therefore, these results should not be extrapolated to all clinical situations and need to be adapted individually to each patient.

The limitations of this study are inherent in the use of observational data, with their higher RoB than randomised clinical trials. The selected studies were all good quality, with a Newcastle-Ottawa Scale score >6. No study on rheumatic diseases was published after 2019, which may be related to the impact of the pandemic on research priorities. All included studies were observational studies, with their risk of confounding by indication (eg, more patients with a history of cancer-initiated rituximab vs other bDMARDs), confounding by other factors, underreporting or misclassification. Although such biases are certainly possible, the results of our meta-analysis are homogenous and do not suggest large variations in different settings. Of note,

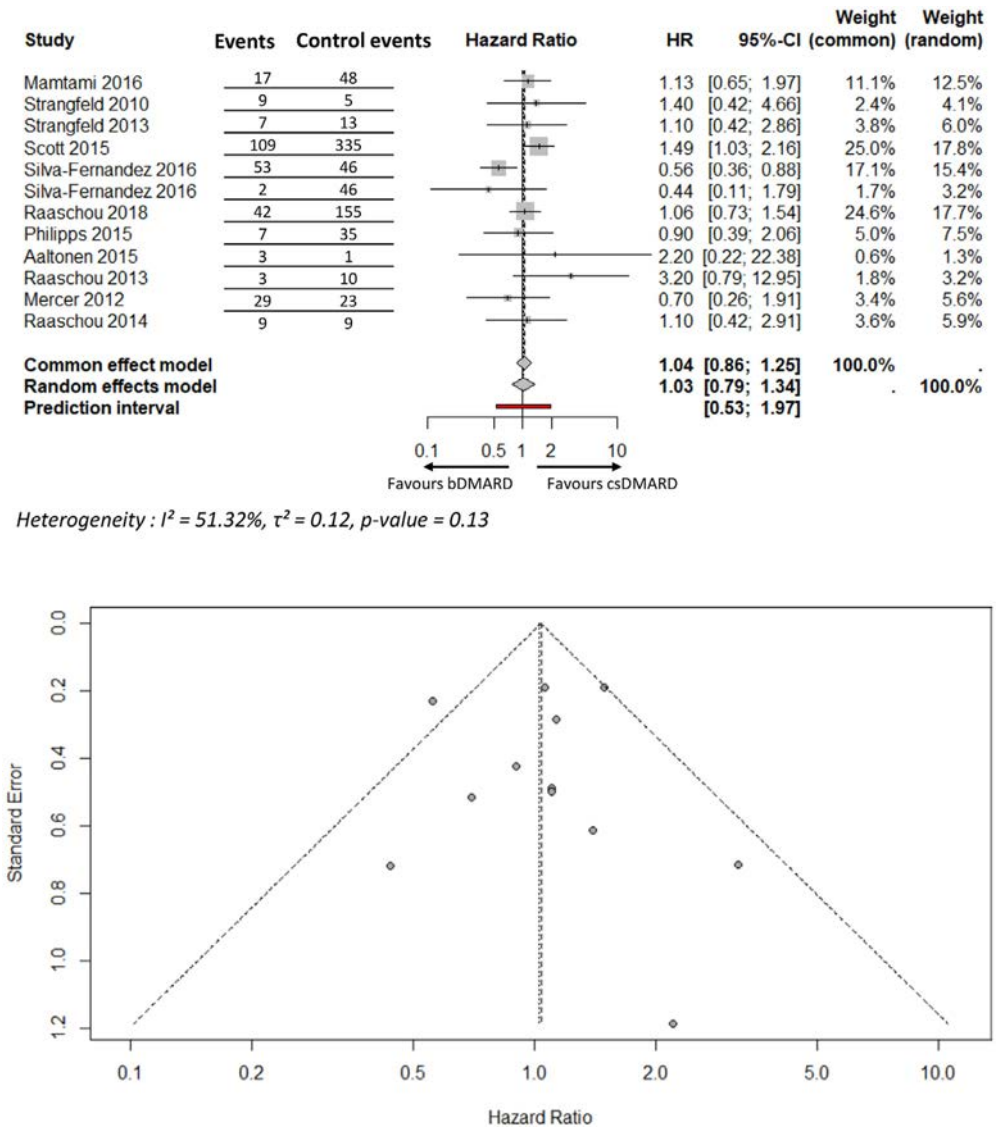


Figure 5. Sensitivity analysis restricted to patients with rheumatoid arthritis. bDMARDs, biologic disease-modifying antirheumatic drugs; csDMARDs, conventional synthetic disease-modifying antirheumatic drugs.

most of the studies adjusted the reported HR on cancer-related prognostic factors such as stage, metastatic status, alcohol consumption and smoking persistence. A few studies focused on a specific cancer, notably lymphomas. Some important data were missing. 12 of 17 studies reported the time from cancer diagnosis to the initiation of a bDMARD. Only 4 of 17 studies reported follow-ups on targeted therapy for longer than 5 years. Furthermore, the duration of the follow-up was limited, considering the risk of cancer recurrence in the long term, ranging from 9.8 months to 8.8 years. Considering the clinical context covered by the SLR and the current limitations of the literature, we emphasise that the absence of evidence of increased risk of new cancer with targeted therapies versus csDMARDs is not evidence of absence of increased risk.

Notwithstanding these limitations, data on TNF inhibitors in patients with a history of cancer are consistent and reassuring. Data in the literature allowed for comparing the risk of new cancer or cancer recurrence with a TNF inhibitor versus csDMARDs based on the time from the cancer diagnosis to the initiation of treatment (<5 or ≥5 years). Of note, we found no significant difference between therapies in new cancer or cancer recurrence, even in patients treated less than 5 years since the diagnosis of cancer. This result is particularly important because some previous guidelines recommended avoiding the use of targeted therapies in patients with a recent cancer [27,28] which, as a consequence, may make any selection away from b/tsDMARDs stronger the shorter the time interval between the cancer and the start of a b/tsDMARD. When evaluating the benefit/risk of initiating or re-initiating a targeted therapy in patients with a history of cancer, prioritising cancer-related risk factors for recurrence might hold greater significance than focusing solely on the time elapsed since the initial cancer diagnosis. In addition, from the collected data, longer exposure to a TNF inhibitor >5 years was not associated with increased risk of a new cancer or cancer recurrence as compared with csDMARD treatment.

Unexpectedly, despite the wide use of this bDMARD in current practice [7], in patients with a history of cancer, limited data were available for rituximab in patients with IA and a history of cancer. This absence of epidemiological data, along with recent translational results showing the role of B lymphocytes in tumour surveillance, should serve as an incentive to generate more data on rituximab in this setting [29]. The limited or lack of data on interleukin 12/23 (IL-12/23), IL-23 and IL-17 inhibitors in patients with a history of cancer should also be added to the collective research agenda. Data on abatacept and Janus kinase/signal regulator and activator of transcription (JAK/

STAT) inhibitors in patients with a history of cancer will be of specific interest, given the results of some studies, out of scope for this SLR because these studies reported data on patients without a history of cancer. In the context of RA without a history of cancer, some observational studies reported an increased risk of cancer with abatacept versus other targeted therapies [30–32]. Of note, the mechanism of action of abatacept is the opposite of that of ipilimumab used in cancer immunotherapy, which warrants some heightened monitoring. JAK/STAT inhibitors have also raised concerns regarding the risk of cancer after the Oral-Surveillance study, which excluded patients with a history of cancer [33].

In summary, this SLR informing the EULAR PTC on the initiation of targeted therapies in patients with IA and a history of cancer shows that overall, the targeted therapies and clinical context covered by the included studies were not associated with increased risk of new cancer or cancer recurrence as compared with csDMARDs. We further found no significant risk by type of cancer, time from cancer to the initiation of a bDMARD and treatment duration.

Results of the present meta-analysis based on studies evaluating the risk of recurrent or new cancer after targeted therapy initiation with a very limited follow-up should be interpreted with caution. Additional data are needed regarding inflammatory rheumatic diseases other than RA, treatments other than TNF inhibitors as well as longer durations of follow-up and exposure to targeted therapies. Improving the data collection, notably on non-TNF targeted therapies, and improving the reporting, differentiating relapses of a previous cancer from new incident cancers, and extending the follow-up of patients are needed in future studies, for patients and clinicians.

Acknowledgments

The Task Force gratefully thanks the librarian Mafalda Burri (Bibliothèque de l'Université de Genève, Genève, Suisse) for her contribution to the systematic literature search, and European League Against Rheumatism for financial and logistical support. The results of this systematic literature review were presented at the 2023 EULAR Congress [35].

Contributors

ES and JM-C wrote the first draft of the manuscript, with help from KL, AF and J-EG. All authors participated in the work of the Task Force, provided coauthor contribution to the manuscript and read and approved the final manuscript. J-EG, as guarantor, accept full responsibility for the work, had access to the data and control the decision to publish.

Funding

This project was funded by European Alliance of Associations for Rheumatology.

Competing interests

Participants provided declarations of interest; the individual declarations will be attached as an online supplemental file.

Patient and public involvement

Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication

Not applicable.

Ethics approval

Not applicable.

Data availability statement

Data are available upon reasonable request. The data underlying this article are available in the article and in the online supplementary material.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1136/ard-2024-225981.

Orcid

Eden Sebbag: http://orcid.org/0000-0003-3045-7725
Juan Molina-Collada: http://orcid.org/0000-0001-5191-7802
Daniel Aletaha: http://orcid.org/0000-0003-2108-0030
Johan Askling: http://orcid.org/0000-0003-0433-0616
Samuel Bitoun: http://orcid.org/0000-0003-0891-2269
Ertugrul Cagri Bolek: http://orcid.org/0000-0003-3886-2813
Maya H. Buch: http://orcid.org/0000-0002-8962-5642
Gerd R. Burmester: http://orcid.org/0000-0001-7518-1131
Katerina Chatzidionysiou: http://orcid.org/0000-0002-2669-1247
Francois-Xavier Danlos: http://orcid.org/0000-0003-1880-064X
Merete Lund Hetland: http://orcid.org/0000-0003-4229-6818
Florenzo Iannone: http://orcid.org/0000-0003-0474-5344
Marie Kostine: http://orcid.org/0000-0002-6729-6200
Tue Wenzel Kragstrup: http://orcid.org/0000-0002-6439-397X
Tore K. Kvien: http://orcid.org/0000-0002-8441-3093
Anne Constanze Regierer: http://orcid.org/0000-0003-2456-4049
Hendrik Schulze-Koops: http://orcid.org/0000-0002-1681-491X
Kim Lauper: http://orcid.org/0000-0002-4315-9009
Axel Finckh: http://orcid.org/0000-0002-1210-4347
Jacques-Eric Gottenberg: http://orcid.org/0000-0002-9469-946X

REFERENCES

[1] Smolen JS, Landewé RBM, Bergstra SA, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis 2023;82:3–18. doi: 10.1136/ard-2022-223356.

[2] Ramiro S, Nikiphorou E, Sepriano A, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. Ann Rheum Dis 2023;82:19–34. doi: 10.1136/ard-2022-223296.
[3] Gossec L, Baraliakos X, Kerschbaumer A, et al. EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2019 update. Ann Rheum Dis 2020;79:700–12. doi: 10.1136/annrheumdis-2020-217159.
[4] Simon TA, Thompson A, Gandhi KK, et al. Incidence of malignancy in adult patients with rheumatoid arthritis: a meta-analysis. Arthritis Res Ther 2015;17:212. doi: 10.1186/s13075-015-0728-9.
[5] Abásolo L, Júdez E, Descalzo MA, et al. Cancer in rheumatoid arthritis: occurrence, mortality, and associated factors in a South European population. Semin Arthritis Rheum 2008;37:388–97. doi: 10.1016/j.semarthrit.2007.08.006.
[6] de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. Nat Rev Cancer 2006;6:24–37. doi: 10.1038/nrc1782.
[7] Chatzidionysiou K, Delcoigne B, Frisell T, et al. How do we use biologics in rheumatoid arthritis patients with a history of malignancy? An assessment of treatment patterns using Scandinavian registers. RMD Open 2020;6:e001363. doi: 10.1136/rmdopen-2020-001363.
[8] Rayyan - AI powered tool for systematic literature reviews 2021. Available: https://www.rayyan.ai/.
[9] Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919. doi: 10.1136/bmj.i4919.
[10] Mamtani R, Clark AS, Scott FI, et al. Association Between Breast Cancer Recurrence and Immunosuppression in Rheumatoid Arthritis and Inflammatory Bowel Disease: A Cohort Study. Arthritis Rheumatol 2016;68:2403–11. doi: 10.1002/art.39738.
[11] Raaschou P, Frisell T, Askling J, et al. TNF inhibitor therapy and risk of breast cancer recurrence in patients with rheumatoid arthritis: a nationwide cohort study. Ann Rheum Dis 2015;74:2137–43. doi: 10.1136/annrheumdis-2014-205745.
[12] Raaschou P, Simard JF, Holmqvist M, et al. Rheumatoid arthritis, anti-tumour necrosis factor therapy, and risk of malignant melanoma: nationwide population based prospective cohort study from Sweden. BMJ 2013;346:bmj.f1939. doi: 10.1136/bmj.f1939.
[13] Scott FI, Mamtani R, Brensinger CM, et al. Risk of Nonmelanoma Skin Cancer Associated With the Use of Immunosuppressant and Biologic Agents in Patients With a History of Autoimmune Disease and Nonmelanoma Skin Cancer. JAMA Dermatol 2016;152:164–72. doi: 10.1001/jamadermatol.2015.3029.
[14] Mercer LK, Green AC, Galloway JB, et al. The influence of anti-TNF therapy upon incidence of keratinocyte skin cancer in patients with rheumatoid arthritis: longitudinal results from the British Society for Rheumatology Biologics Register. Ann Rheum Dis 2012;71:869–74. doi: 10.1136/annrheumdis-2011-200622.
[15] Phillips C, Zeringue AL, McDonald JR, et al. Tumor Necrosis Factor Inhibition and Head and Neck Cancer Recurrence and Death in Rheumatoid Arthritis. PLoS ONE 2015;10:e0143286. doi: 10.1371/journal.pone.0143286.
[16] Waljee AK, Higgins PDR, Jensen CB, et al. Anti-tumour necrosis factor-α therapy and recurrent or new primary cancers in patients with inflammatory bowel disease, rheumatoid arthritis, or psoriasis and previous cancer in Denmark: a nationwide, population-based cohort study. Lancet Gastroenterol Hepatol 2020;5:276–84. doi: 10.1016/S2468-1253(19)30362-0.
[17] Raaschou P, Söderling J, Turesson C, et al. Tumor Necrosis Factor Inhibitors and Cancer Recurrence in Swedish Patients With Rheumatoid Arthritis: A Nationwide Population-Based Cohort Study. Ann Intern Med 2018;169:291–9. doi: 10.7326/M17-2812.
[18] Aaltonen KJ, Joensuu JT, Virkki L, et al. Rates of serious infections and malignancies among patients with rheumatoid arthritis receiving either tumor necrosis factor inhibitor or rituximab therapy. J Rheumatol 2015;42:372–8. doi: 10.3899/jrheum.140853.
[19] Strangfeld A, Hierse F, Rau R, et al. Risk of incident or recurrent malignancies among patients with rheumatoid arthritis exposed to biologic therapy in the German biologics register RABBIT. Arthritis Res Ther 2010;12:R5. doi: 10.1186/ar2904.
[20] Axelrad J, Bernheim O, Colombel J-F, et al. Risk of New or Recurrent Cancer in Patients With Inflammatory Bowel Disease and Previous Cancer Exposed to Immunosuppressive and Anti-Tumor Necrosis Factor Agents. Clin Gastroenterol Hepatol 2016;14:58–64. doi: 10.1016/j.cgh.2015.07.037.
[21] Vedamurthy A, Gangasani N, Ananthakrishnan AN. Vedolizumab or Tumor Necrosis Factor Antagonist Use and Risk of New or Recurrent Cancer in Patients With Inflammatory Bowel Disease With Prior Malignancy: A Retrospective Cohort Study. Clin Gastroenterol Hepatol 2022;20:88–95. doi: 10.1016/j.cgh.2020.10.007.
[22] Risk of cancer recurrence or new tumors in RA patients with prior malignancies treated with various biologic agents. ACR Meeting Abstracts Available: https://acrabstracts.org/abstract/risk-of-cancer-recurrence-or-new-tumors-in-ra-patients-with-prior-malignancies-treated-with-various-biologic-agents/.
[23] Silva-Fernández L, Lunt M, Kearsley-Fleet L, et al. The incidence of cancer in patients with rheumatoid arthritis and a prior malignancy who receive TNF inhibitors or rituximab: results from the British Society for Rheumatology Biologics Register-Rheumatoid Arthritis. Rheumatology (Sunnyvale) 2016;55:2033–9. doi: 10.1093/rheumatology/kew314.
[24] Yan MK, Wang C, Wolfe R, et al. Association Between Low-Dose Methotrexate Exposure and Melanoma: A Systematic Review and Meta-analysis. JAMA Dermatol 2022;158:1157–66. doi: 10.1001/jamadermatol.2022.3337.
[25] Davis JM. Overview of the Associations Between Cancer and Rheumatic Disease. Rheum Dis Clin North Am 2020;46:417–27. doi: 10.1016/j.rdc.2020.05.002.
[26] De Cock D, Hyrich K. Malignancy and rheumatoid arthritis: Epidemiology, risk factors and management. Best Pract Res Clin Rheumatol 2018;32:869–86. doi: 10.1016/j.berh.2019.03.011.
[27] Goëb V, Ardizzone M, Arnaud L, et al. Recommendations for using TNFα antagonists and French Clinical Practice Guidelines endorsed by the French National Authority for Health. Joint Bone Spine 2013;80:574–81. doi: 10.1016/j.jbspin.2013.09.001.
[28] Rahier JF, Magro F, Abreu C, et al. Second European evidence-based consensus on the prevention, diagnosis and management of opportunistic infections in inflammatory bowel disease. J Crohns Colitis 2014;8:443–68. doi: 10.1016/j.crohns.2013.12.013.
[29] Fridman WH, Meylan M, Petitprez F, et al. B cells and tertiary lymphoid structures as determinants of tumour immune contexture and clinical outcome. Nat Rev Clin Oncol 2022;19:441–57. doi: 10.1038/s41571-022-00619-z.
[30] Ozen G, Pedro S, Schumacher R, et al. Safety of abatacept compared with other biologic and conventional synthetic disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: data from an observational study. Arthritis Res Ther 2019;21:141. doi: 10.1186/s13075-019-1921-z.
[31] Montastruc F, Renoux C, Dell'Aniello S, et al. Abatacept initiation in rheumatoid arthritis and the risk of cancer: a population-based comparative cohort study. Rheumatology (Sunnyvale) 2019;58:683–91. doi: 10.1093/rheumatology/key352.
[32] Simon TA, Boers M, Hochberg M, et al. Comparative risk of malignancies and infections in patients with rheumatoid arthritis initiating abatacept versus other biologics: a multi-database real-world study. Arthritis Res Ther 2019;21:228. doi: 10.1186/s13075-019-1992-x.
[33] Curtis JR, Yamaoka K, Chen Y-H, et al. Malignancy risk with tofacitinib versus TNF inhibitors in rheumatoid arthritis: results from the open-label, randomised controlled ORAL Surveillance trial. Ann Rheum Dis 2023;82:331–43. doi: 10.1136/ard-2022-222543.
[34] Hasan B, Tandon KS, Miret R, et al. Ustekinumab does not increase risk of new or recurrent cancer in inflammatory bowel disease patients with prior malignancy. J Gastroenterol Hepatol 2022;37:1016–21. doi: 10.1111/jgh.15806.
[35] Sebbag E, Molina Collada J, Lauper K, et al. POS1069 SYSTEMATIC literature review informing the EULAR points to consider task force on the initiation of targeted therapies in patients with inflammatory arthritides and a history of cancer. Ann Rheum Dis 2023;82:856.

Ann Rheum Dis 84 (2025) 653–656

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Letter

Teclistamab in relapsed systemic sclerosis after autologous haematopoietic stem cell transplantation

To the editor,

Systemic sclerosis (SSc) is a devastating autoimmune disease characterised by tissue fibrosis and obliterative vasculopathy. Presence of specific autoantibodies suggests a central role of B-cells and plasma cells (PCs), which are enriched in affected skin lesions [1] and might contribute to tissue fibrosis, with LGR5-expressing fibroblasts playing a major role [2]. CD19 chimeric antigen receptor (CAR)-T cell therapy has recently provided a novel therapeutic approach in refractory SSc through deep B-cell depletion [3], but its widespread use is limited by the expensive and time-consuming production process and the requirement of conditioning-regimens. Bispecific antibodies may represent an alternative therapeutic option given their high potency in target cell killing. They have dual binding sites directed against CD3 and a target cell antigen, such as CD19 or B-cell maturation antigen (BCMA), resulting in $CD3^+$ T-cell activation and subsequent target cell killing. Their use in SSc recently provided remarkable responses in 2 reported cases [4,5].

With this background, we utilised teclistamab, a CD3 and BCMA-bispecific antibody, approved for the treatment of multiple myeloma [6], in a 48-year-old female patient with severe SSc. She showed disease progression with worsening skin involvement and interstitial lung disease (ILD) despite previous treatments with azathioprine, cyclophosphamide, mycophenolate mofetil (MMF), and nintedanib. After an initial response to autologous haematopoietic stem cell transplantation (HSCT), she relapsed 12 months later with progressive diffuse skin disease, disabling contractures, ILD, and new-onset myocardial involvement. Due to rapid disease progression refractory to combination therapy with rituximab (RTX) and MMF and an American College of Rheumatology Composite Response Index in Systemic Sclerosis (ACR-CRISS) score of 0% probability of improvement, we decided to treat her with teclistamab. She received a step-up dosage of subcutaneous teclistamab at 0.06, 0.3 and 0.8 mg/kg before receiving the full dose of 1.5 mg/kg at weeks 2, 3 and 4. Treatment with MMF was discontinued and peripheral blood B-cells were undetectable following the previous administration of 2000 mg of RTX 5 months prior to teclistamab. Adverse events observed included a grade 1 cytokine release syndrome (Common Terminology Criteria for Adverse Events [CTCAE] grade 1), hypogammaglobulinemia (CTCAE grade 2), gastroenteritis (CTCAE grade 1), and common cold (CTCAE grade 2); no serious adverse events were observed.

After teclistamab treatment, skin fibrosis rapidly improved with a reduction of the modified Rodnan skin score from 36 at baseline to 21 at week 20 (Fig, A). In addition, her lung function gradually improved with normalisation of forced vital capacity at week 16 and regression of morphologic signs of pulmonary inflammation and fibrosis on chest computed tomography (CT) and fibroblast-activated protein (FAPI)-positron emission tomography (PET) (Fig, A; Supplementary Fig S1). Myocarditis also improved, reflected by normalisation of troponin T levels and of left ventricular ejection fraction, along with a reduction of gadolinium late enhancement on cardiac magnetic resonance imaging and decreased fibroblast activation protein (FAP) activity on PET imaging of the myocardial septum (Fig, A). Furthermore, we observed a recovery of acral perfusion following a standardised cold-water challenge (Supplementary Fig S2). Cumulatively, these changes resulted in a significant improvement in health-related quality of life (Supplementary Fig S3) and reduction in disease activity indices (Supplementary Fig S4). The beneficial clinical responses were accompanied by a reduction of antinuclear and anti-Scl-70 antibodies, both of which nearly normalised (Fig, B). In parallel, serum immunoglobulin G (IgG) and vaccine titres also declined (Fig, B; Supplementary Fig S5), suggesting a depletion of large fractions of autoreactive as well as of protective PCs. Indeed, PCs and B-cells were completely eradicated from the bone marrow at week 8 (Supplementary Fig S6), and largely depleted in affected skin (Fig, C). Skin biopsies also showed an abrogation of FAP expression and disappearance of LGR5-expressing fibroblasts, as well as revascularisation and a recovery of vascular architecture (Fig, C).

In conclusion, we demonstrate rapid and sustained improvement across all clinical SSc disease domains in response to teclistamab treatment, which was accompanied by deep tissue depletion of B-cells and PCs, as well as a decrease in hyperactive fibroblasts as the major hallmark of SSc. These data provide a first signal that teclistamab is a promising therapeutic option, provided that larger prospective and controlled studies confirm these findings and that immunological side effects do not pose unacceptable risks. Specific caution should be given to immunoglobulin levels and vaccine titres, with IVIG use and revaccination carefully evaluated.

ES, RB, GK, and TA contributed equally.

https://doi.org/10.1016/j.ard.2025.01.043
Received 25 November 2024; Revised 21 January 2025; Accepted 25 January 2025

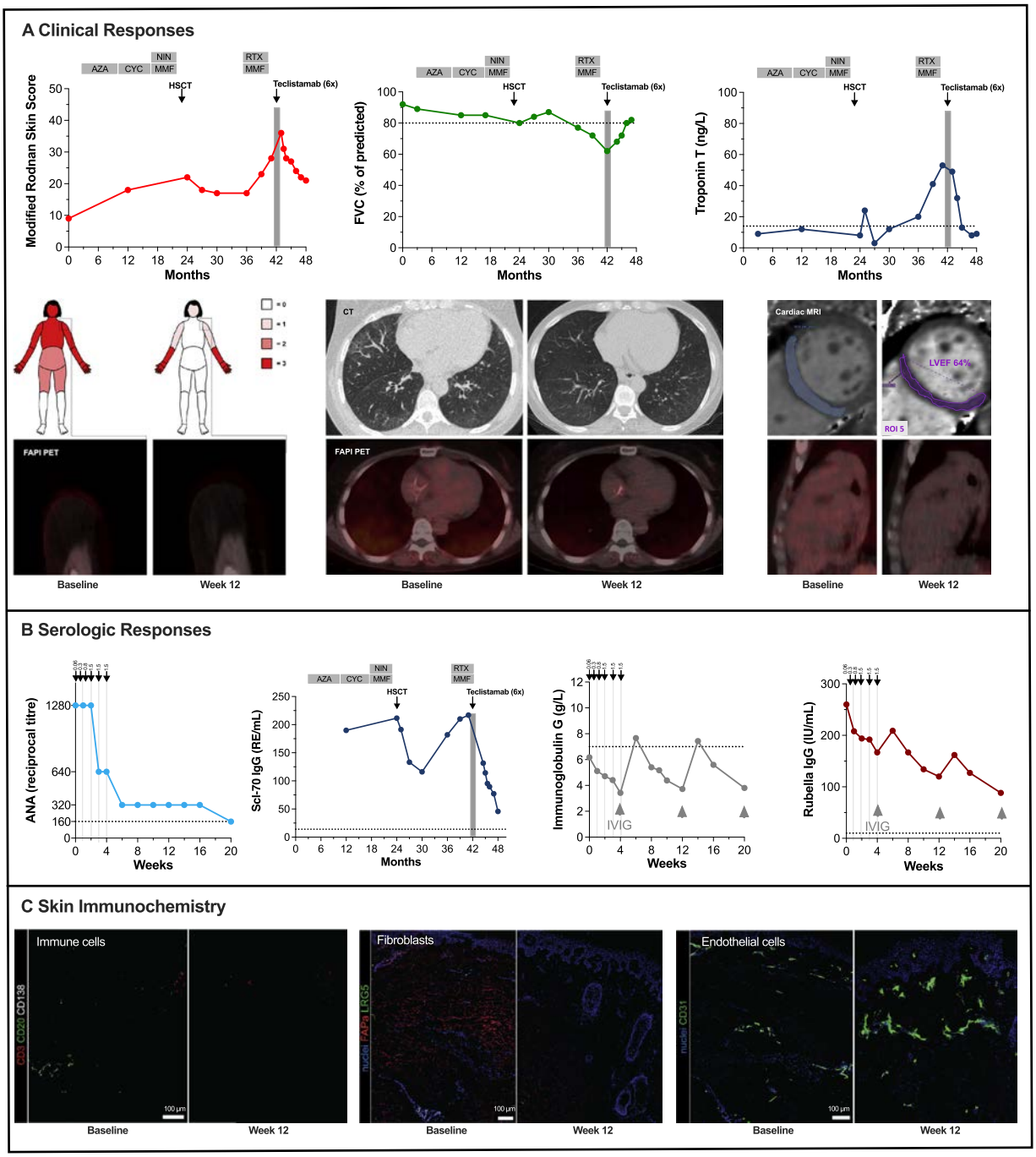


Figure. (A) The modified Rodnan skin score (upper left), forced vital capacity (middle), and troponin levels in blood (right) in relation to the patient's medication history. HSCT refers to autologous haematopoietic stem cell transplantation after conditioning with 200 mg/kg cyclophosphamide and 23.5 mg/kg rabbit anti-thymocyte globulin (Grafalon). Teclistamab (× 6) refers to the 6 doses that were administered. Below are representative images from a fibroblast-activated protein (FAPI)-positron emission tomography (PET) performed at baseline and week 12 in the affected skin (right upper arm), lung and heart, complemented by computed tomography (CT) scans of the lung (middle) and cardiac magnetic resonance imaging (MRI) (right) that revealed an improvement in left ventricular ejection fraction and reduction of myocardial tracer uptake in the myocardium of the septum. On FAPI-PET of the lungs we saw a reduction of standardised uptake value (SUV)peak (right lung 2.89 to 2.65; left lung 2.44 to 2.47), SUVmean (right lung 2.5 to 2.2; left lung 2.2 to 1.9) and MAV (right lung 259.20 ccm to 223.10 ccm; left lung 259.20 ccm to 204.67 ccm). On FAPI-PET of the myocardial septum, the SUVpeak decreased from 1.58 at baseline to 1.28 at follow-up and the SUVmean from 1.29 to 1.00, respectively. (B) Changes in the serum levels of antinuclear antibodies (ANA), anti-Scl-70 antibodies, immunoglobulin G, and vaccine titres for rubella. The dotted lines indicate the upper limit of the normal range for ANA and anti-Scl-70 and the lower limit of normal for IgG and vaccine titres. (C) Representative results of immunohistochemistry from skin biopsies of affected skin (left lateral upper arm) at baseline and week 12 with depletion of tissue-infiltrating $CD20^+$ B-cells and $CD138^+$ plasma cells (left), reduction of FAP-alpha and LRG5-expression, 2 markers expressed by fibroblasts (middle), and $CD31^+$ endothelial cells (right), demonstrating improvement of the initially damaged microvasculature.

Competing interests

JH received speakers fee or honoraria from Janssen-Cilag. JK received honoraria and support for attending meetings from Johnson & Johnson. DS received grants or contracts from Johnson & Johnson and consulting fees from Janssen-Cilag. AK received consulting fees from Janssen Pharmaceutica. UK received travel support, honoraria for advisory boards and

invited lectures from Johnson & Johnson/Janssen Germany. GK received honoraria for advisory boards and invited lectures from Johnson & Johnson. TA received a study grant from Johnson & Johnson. ES, RB, MD, CF, SF, MP, FD, TWR, AF, AW, AB, EW, ACP, AK, and GRB declared no conflict of interest in relation to this manuscript.

Acknowledgements

We thank Skander Elleuche from EUROIMMUNE AG for the assessment of anti-topoisomerase antibodies and Klaus Engel from Siemens Healthineers for cinematic rendering of the FAPI-PET images.

Funding

Supported by a grant from the European Union (Horizon 2020 ERC-2020-CoG 101001866-INSPIRE) to Dr. Gerhard Krönke.

Patient consent for publication

Not applicable

Ethics approval

Not applicable

Provenance and peer review

Not applicable

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ard.2025.01.043.

Orcid

Robert Biesen: http://orcid.org/0000-0002-0434-7832
Maria Dzamukova: http://orcid.org/0000-0002-2389-3222
Meike Probst: http://orcid.org/0009-0004-3390-6789
Felix Doellinger: http://orcid.org/0000-0003-3661-8569
Ann-Christin Pecher: http://orcid.org/0000-0003-0201-0686
Jörg Henes: http://orcid.org/0000-0002-8385-6861
Tobias Alexander: http://orcid.org/0000-0003-1193-0097

REFERENCES

[1] Whitfield ML, Finlay DR, Murray JI, Troyanskaya OG, Chi JT, Pergamenschikov A, et al. Systemic and cell type-specific gene expression patterns in scleroderma skin. Proc Natl Acad Sci U S A. 2003;100(21):12319–24.
[2] Gur C, Wang SY, Sheban F, Zada M, Li B, Kharouf F, et al. LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma. Cell 2022;185(8):1373–1388.e1320.
[3] Auth J, Müller F, Völkl S, Bayerl N, Distler JHW, Tur C, et al. CD19-targeting CAR T-cell therapy in patients with diffuse systemic sclerosis: a case series. Lancet Rheumatol 2025;7(2):e83–93.
[4] Subklewe M, Magno G, Gebhardt C, Bücklein V, Szelinski F, Arévalo HJR, et al. Application of blinatumomab, a bispecific anti-CD3/CD19 T-cell engager, in treating severe systemic sclerosis: A case study. Eur J Cancer 2024;204:114071.
[5] Hagen M, Bucci L, Böltz S, Nöthling DM, Rothe T, Anoshkin K, et al. BCMA-targeted T-cell-engager therapy for autoimmune disease. N Engl J Med 2024;391(9):867–9.
[6] Moreau P, Garfall AL, van de Donk N, Nahi H, San-Miguel JF, Oriol A, et al. Teclistamab in relapsed or refractory multiple myeloma. N Engl J Med 2022;387(6):495–505.

Elise Siegert, Robert Biesen
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Maria Dzamukova
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
German Rheumatology Research Centre (DRFZ) Berlin, a Leibniz Institute, Berlin, Germany

Christian Furth
Department of Nuclear Medicine, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Meike Probst
Department of Infectious Diseases, Respiratory and Critical Care Medicine, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Felix Doellinger, Thula Walter-Rittel
Department of Radiology, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Anja Fleischmann, Artur Wilhelm, Anne Elisabeth Beenken
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
German Rheumatology Research Centre (DRFZ) Berlin, a Leibniz Institute, Berlin, Germany

Edgar Wiebe
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Ann-Christin Pecher, Jörg Henes
Center for Interdisciplinary Rheumatology, Immunology and Autoimmune Diseases (INDIRA) and Department for Internal Medicine II (Hematology, Oncology, Rheumatology and Immunology), Universitätsklinikum Tübingen, Tübingen, Germany

Stefan Florian
Institute of Pathology, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

David Simon, Arnd Kleyer
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

Gerd-Rüdiger Burmester
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
German Rheumatology Research Centre (DRFZ) Berlin, a Leibniz Institute, Berlin, Germany

Ulrich Keller, Jan Krönke
Department of Hematology, Oncology and Cancer Immunology, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
German Center for Translational Cancer Research (DKTK), Partner Site Berlin, Berlin, Germany

Gerhard Krönke, Tobias Alexander*
Department of Rheumatology and Clinical Immunology, Charité – Universitätsmedizin Berlin Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany
German Rheumatology Research Centre (DRFZ) Berlin, a Leibniz Institute, Berlin, Germany
***Correspondence to** Dr Tobias Alexander, Department of Rheumatology and Clinical Immunology, Charitéplatz 1, Charité, Universitätsmedizin Berlin, Germany
E-mail address: tobias.alexander@charite.de (T. Alexander).

Ann Rheum Dis 84 (2025) e19–e20

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Correspondence

Why the term nr-axSpA should not be used for diagnosis

Dear Sir,

We read with great interest the recent impressive immunohistologic description of a patient with axial spondyloarthritis (axSpA) and cardiac valve involvement [1]. Inflammation at this site is rare but a well-known feature of radiographic axSpA (r-axSpA), which is equivalent to ankylosing spondylitis (AS) [2]. In this report on a female patient with axSpA for 24 years, the authors stress that they have "*diagnosed*" nonradiographic (nr) axSpA [1]. However, if this classification is important ("*first report*"), the absence of erosions in the sacroiliac joints (SIJs) and the degree of axial inflammation and bone formation should also be reported. It is necessary to have an estimate of the musculoskeletal burden of the disease of the patient, which is more important than a rather artificial classification based on the assumption of a fundamental difference between the 2 subtypes.

The recent Assessments in axial SpondyloArthritis international Society statement contained expert agreement on axSpA as being the overall term for the disease [2]. The use of the term nr-axSpA for diagnosis does not make sense for different reasons. Most important is the known difficulty in quantifying radiographic changes in the SIJs [3] since the cutoff between r-axSpA and nr-axSpA (<grade 2 according to the 1984 modified New York [mNY] criteria) is too subjective, and the differentiation is not reliable. Furthermore, the prevalence of erosions in the SIJs, as detected by magnetic resonance imaging (MRI), was not much different between subtypes. Conventional radiography is likely to be soon replaced by MRI as the main tool to assess not only active but also structural changes in the SIJs [4]. The 2 subtypes are clearly part of one disease, even though there are some differences. The favourable outcomes reported in recent cohort studies may be due to the time point at which patients were included [5].

The main reason for the introduction and use of the term nr-axSpA more than a decade ago was the historical situation in which biologic disease modifying antirheumatic drug approved for active AS (fulfilling mNY criteria) could not be used for early patients. In contrast, the 2009 Assessments in axial SpondyloArthritis international Society classification criteria enabled the classification of axSpA patients without radiographic changes in the SIJs. Thus, based on the need to obtain approval for early disease, the term nr-axSpA was established. In trials on nr-axSpA, authorities perform quality checks on the degree of radiographic SIJ changes. This has and may still lead to the exclusion of patients differently scored by central readers (patients with r-axSpA are not allowed in the study) as a matter of classification.

Furthermore, nr-axSpA patients may well have syndesmophytes; the prevalence was almost 10% in a cohort study, and in another cohort study, 24% of patients without sacroiliitis had evidence of spinal inflammation. This mixture makes it hardly possible for a clinician to use the term nr-axSpA for diagnosis.

The main thought behind putting SIJ changes so much in the foreground is based on the mNY criteria, which served well for some decades because the vast majority of AS patients recognised at that time did have radiographic SIJ changes. However, the functional relevance of such changes was shown to be limited [6]. For axSpA patients, the most important outcome, next to the level of pain and stiffness, is the degree of spinal involvement —first by inflammation and later by syndesmophytes and ankylosis, both strongly associated with function and mobility. The latter is due to bone formation, which is usually irreversible.

Taken together, the term nr-axSpA should not be used for diagnosis [2]. Nevertheless, the involvement of the heart in axSpA is a very interesting special feature of the disease that needs more sophisticated research—such as provided by the authors [1].

Competing interests

All authors declare they have no competing interests.

Funding

There was no funding for this letter.

Patient consent for publication

There was no public or patient involvement.

Ethics approval

Not applicable.

Provenance and peer review

Not commissioned; externally peer reviewed.

DOI of original article: http://dx.doi.org/10.1016/j.ard.2024.12.003.

Handling editor Josef S. Smolen.

https://doi.org/10.1016/j.ard.2024.11.003

Received 13 November 2024; Accepted 16 November 2024

Declaration of generative AI and AI-assisted technologies in the writing process

There was no artificial intelligence involved.

Orcid

Juergen Braun: http://orcid.org/0000-0002-9156-5095

REFERENCES

[1] Qaiyum Z, Seidman MA, Ross HJ, Inman RD. Cardiac valve immune profiling in non-radiographic axial spondyloarthritis. Ann Rheum Dis 2024;83(10):1394–9.

[2] van der Heijde D, Molto A, Ramiro S, Braun J, Dougados M, van Gaalen FA, et al. Goodbye to the term 'ankylosing spondylitis', hello 'axial spondyloarthritis': time to embrace the ASAS-defined nomenclature. Ann Rheum Dis 2024;83(5):547–9.

[3] van Tubergen A, Heuft-Dorenbosch L, Schulpen G, Landewé R, Wijers R, van der Heijde D, et al. Radiographic assessment of sacroiliitis by radiologists and rheumatologists: does training improve quality? Ann Rheum Dis 2003;62(6):519–25.

[4] Poddubnyy D, Diekhoff T, Baraliakos X, Hermann KGA, Sieper J. Diagnostic evaluation of the sacroiliac joints for axial spondyloarthritis: should MRI replace radiography? Ann Rheum Dis 2022;81(11):1486–90.

[5] Braun J. The early course of radiographic progression in axial spondyloarthritis: have we missed something? J Rheumatol 2024 jrheum.2024-0807. doi: 10.3899/jrheum.2024-0807.

[6] Poddubnyy D, Listing J, Haibel H, Knüppel S, Rudwaleit M, Sieper J. Functional relevance of radiographic spinal progression in axial spondyloarthritis: results from the GErman SPondyloarthritis Inception Cohort. Rheumatology (Oxford) 2018;57(4):703–11.

Juergen Braun *

Rheumatologisches Versorgungszentrum Steglitz, Berlin, Germany

Ruhr University Bochum, Germany

*__Correspondence to__ Prof Juergen Braun.

E-mail address: juebraun@gmx.de

Ann Rheum Dis 84 (2025) e21–e22

Contents lists available at ScienceDirect

Annals of the Rheumatic Diseases

journal homepage: https://www.sciencedirect.com/journal/annals-of-the-rheumatic-diseases

Response

Response to Correspondence on 'Why the term nr-axSpA should not be used for diagnosis'

We were pleased to receive Dr Braun's thoughtful comments on our publication describing the cell biology of cardiac valvulitis in a patient with nonradiographic axial spondyloarthritis [1]. Dr Braun raises some timely comments on nomenclature in this field because it is a changing landscape. A recent publication proposes that the term ankylosing spondylitis be supplanted by axial spondyloarthritis in future publications. We are in full agreement with that, and indeed one of us (R.I.) is a coauthor on that proposal [2]. In the clinical sphere, the domain of axSpA has been stratified into those with diagnostic radiographs (radiographic axSpA [r-axSpA]) and those with other confirmatory tests, most commonly magnetic resonance imaging (MRI) (nonradiographic axSpA [nr-axSpA]). In our case, pelvic radiographs did not fulfill the modified New York criteria for r-axSpA, nor were there any syndesmophytes on spinal X-rays. The MRI documented subchondral bone marrow edema in the sacroiliac joints, without inflammatory lesions detected in the spine.

It should also be acknowledged that the tendency to conflate classification criteria with diagnostic criteria has added another dimension to these discussions [3].

From the early recognition of these categories, there has been discussion whether these clinical subsets constitute distinct diseases or a different expression of a single disease entity [4]. Our report reports the occurrence of aortic valve inflammation in nr-axSpA, which has previously been reported only in r-axSpA. This provides further supportive evidence that both clinical subsets are part of a common disease entity, namely, axSpA. Furthermore, our immune profiling identifies cellular components in the inflamed cardiac valves, which recapitulates those which we have identified in the joints of r-axSpA patients [5].

Acknowledging biologic commonalities between r-axSpA and nr-axSpA, is there any clinical utility in splitting rather than lumping? One of the consequences of splitting in the first instance was that it provided clinicians with the opportunity to initiate biologic therapies in axSpA patients who did not fulfill the modified NY criteria for r-axSpA. This also led to randomized clinical trials in nr-axSpA, which subsequently demonstrated in these patients the effectiveness and safety of tumor necrosis factor (TNF) inhibitors [6,7], interleukin (IL)-17 inhibitors [8], and Jak inhibitors [9,10]. Second, the subdivisions have revealed some differential response patterns when r-axSpA is compared with nr-axSpA, as we have recently reported [11,12].

Our study demonstrates the value of imaging mass cytometry in dissecting the cell biology underlying target tissues in axSpA. The finding of integrin-expressing cells in the cardiac valve implicates the possibility of gut-heart trafficking in this lesion, by analogy with gut-joint trafficking [13]. Furthermore, the spatial immune profiling documented proximity of CD8+ T cells and macrophages support the notion of local antigen presentation to T cells, as has also been supported in studies of r-axSpA by the oligoclonality of T-cell receptor profiling [14]. Further studies incorporating advanced spatial transcriptomics and proteomics hold the promise of future insights in to the pathogenesis of axSpA.

Competing interests

All authors declare they have no competing interests.

Orcid

Robert D. Inman: http://orcid.org/0000-0002-4750-1422

REFERENCES

[1] Qaiyum Z, Seidman MA, Ross HJ, Inman RD. Cardiac valve immune profiling in non-radiographic axial spondyloarthritis. Ann Rheum Dis 2024 ard-2024-225942. doi: 10.1136/ard-2024-225942.

[2] van der Heijde D, Molto A, Ramiro S, Braun J, Dougados M, van Gaalen FA, et al. Goodbye to the term 'ankylosing spondylitis', hello 'axial spondyloarthritis': time to embrace the ASAS-defined nomenclature. Ann Rheum Dis 2024;83(5):547–9. doi: 10.1136/ard-2023-225185.

[3] Aggarwal R, Ringold S, Khanna D, Neogi T, Johnson SR, Miller A, et al. Distinctions between diagnostic and classification criteria? Arthritis Care Res (Hoboken) 2015;67(7):891–7. doi: 10.1002/acr.22583.

[4] Wallis D, Haroon N, Ayearst R, Carty A, Inman RD. Ankylosing spondylitis and nonradiographic axial spondyloarthritis: part of a common spectrum or distinct diseases? J Rheumatol 2013;40(12):2038–41. doi: 10.3899/jrheum.130588.

[5] Qaiyum Z, Gracey E, Yao Y, Inman RD. Integrin and transcriptomic profiles identify a distinctive synovial CD8+ T cell subpopulation in spondyloarthritis. Ann Rheum Dis 2019;78(11):1566–75. doi: 10.1136/annrheumdis-2019-215349.

[6] Callhoff J, Sieper J, Weiß A, Zink A, Listing J. Efficacy of TNFα blockers in patients with ankylosing spondylitis and non-radiographic axial spondyloarthritis: a meta-analysis. Ann Rheum Dis 2015;74(6):1241–8. doi: 10.1136/annrheumdis-2014-205322.

[7] Landewé R, Sieper J, Mease P, Inman RD, Lambert RG, Deodhar A, et al. Efficacy and safety of continuing versus withdrawing adalimumab therapy in maintaining remission in patients with non-radiographic axial spondyloarthritis (ABILITY-3): a multicentre, randomised, double-blind study. Lancet 2018;392(10142):134–44. doi: 10.1016/S0140- 6736(18)31362-X.

https://doi.org/10.1016/j.ard.2024.12.003
Received 5 December 2024; Accepted 9 December 2024

[8] Braun J, Blanco R, Marzo-Ortega H, Gensler LS, Van den Bosch F, Hall S, et al. Two-year imaging outcomes from a phase 3 randomized trial of secukinumab in patients with non-radiographic axial spondyloarthritis. Arthritis Res Ther 2023;25(1):80. doi: 10.1186/s13075-023-03051-5.

[9] Deodhar A, Van den Bosch F, Poddubnyy D, Maksymowych WP, van der Heijde D, Kim TH, et al. Upadacitinib for the treatment of active non-radiographic axial spondyloarthritis (SELECT-AXIS 2): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2022;400(10349):369–79. doi: 10.1016/S0140-6736(22)01212-0.

[10] van den Bosch F, Deodhar A, Poddubnyy D, Maksymowych WP, van der Heijde D, Kim TH, et al. Upadacitinib in active non-radiographic axial spondyloarthritis: 1-year data from a double-blind, randomized, placebo-controlled, phase 3 trial. ACR Open Rheumatol 2024;6(8):470–80. doi: 10.1002/acr2.11669.

[11] Poddubnyy D, Sieper J, Akar S, Muñoz-Fernández S, Haibel H, Diekhoff T, et al. Females with non-radiographic but not with radiographic axial spondyloarthritis are less likely to achieve inactive disease state than males: results from a multicountry prospective observational study [abstract]. Arthritis Rheumatol 2022;74(Suppl 9).

[12] Remalante-Rayco P, Baja E, Baskurt Z, Chim T, Inman R, Dans L, et al. Synergistic impact of the non-radiographic axial spondyloarthritis subtype and female sex on first-line biologic discontinuation: a study of interaction effects. Arthritis Rheumatol 2024;76(Suppl 9).

[13] Qaiyum Z, Lim M, Inman RD. The gut-joint axis in spondyloarthritis: immunological, microbial, and clinical insights. Semin Immunopathol 2021;43(2):173–92. doi: 10.1007/s00281-021-00845-0.

[14] Faham M, Carlton V, Moorhead M, Zheng J, Klinger M, Pepin F, et al. Discovery of T cell receptor β motifs specific to hla-b27-positive ankylosing spondylitis by deep repertoire sequence analysis. Arthritis Rheumatol 2017;69(4):774–84. doi: 10.1002/art.40028.

Zoya Qaiyum, Robert D. Inman *

Schroeder Arthritis Institute, University Health Network, Toronto, ON, Canada

***Correspondence to** Dr Robert D. Inman, Schroeder Arthritis Institute, University Health Network, University of Toronto, Toronto, ON, Canada

E-mail address: robert.inman@uhn.ca (R.D. Inman).