Arthritis Care & Research

Aims and Scope

Arthritis Care & Research is an official journal of the American College of Rheumatology and the Association of Rheumatology Professionals, a division of the College. *Arthritis Care & Research* is a peer-reviewed journal that publishes both original research and review articles that promote excellence in the clinical practice of rheumatology. Relevant to the care of individuals with arthritis and related disorders, major topics are evidence-based practice studies, clinical problems, practice guidelines, health care economics, health care policy, educational, social, and public health issues, and future trends in rheumatology practice.

Volume 78 No. 1 January 2026

Arthritis Care & Research

An Official Journal of the American College of Rheumatology
www.arthritiscareres.org and wileyonlinelibrary.com

Arthritis Care & Research

An Official Journal of the American College of Rheumatology
www.arthritiscareres.org and wileyonlinelibrary.com

VOLUME 78 • JANUARY 2026 • NO. 1

Theme Issue: Pain Science and Therapy in Rheumatic Disease

Cover image: The image on the cover (Image made with BioRender; from Kelleher et al; pages 3–14) shows top-down and bottom-up mechanisms of persistent pain in inflammatory arthritis.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 1–2
DOI 10.1002/acr.70025

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Introduction to the Special Theme Issue: Pain Science and Therapy in Rheumatic Disease

Afton L. Hassett,[1] S. Sam Lim,[2] and Kelli D. Allen[3,4]

In this issue of *Arthritis Care & Research* (*AC&R*), we highlight articles focused on pain science and therapy in rheumatic disease. This theme was selected because pain remains one of the most disabling and undertreated symptoms across rheumatic diseases, directly influencing quality of life, physical function, and health disparities. In addition, there are many ongoing questions regarding pain mechanisms and treatment approaches. Manuscripts submitted for themed issues of *AC&R* undergo the same peer-review procedures as other scientific manuscripts in the journal and, therefore, meet the journal's rigorous standards. We are pleased to present in this issue 16 articles spanning a diverse array of topics related to pain science and therapy in rheumatic disease. We would like to highlight some key themes that emerged in this issue.

First, multiple articles addressed pain mechanisms, with an emphasis on nociplastic pain (ie, pain arising from changes in how the central nervous system processes pain signals, often in the absence of tissue damage or inflammation). Overall, these studies suggest a key role for altered central pain processing in rheumatic diseases. Kelleher and colleagues provided a review of studies on brain mechanisms of pain in inflammatory arthritis (IA), concluding that evidence supports a dual "top-down" (eg, pre-existing dysfunction in pain perception) and "bottom-up" (eg, peripheral inflammation acting on the brain) model of persistent pain in IA.[1] One study in this issue found that there was evidence of centrally mediated pain at the time of an IA diagnosis, challenging the notion that at early stages of disease, pain is driven only by peripheral mechanisms.[2] Another study evaluating individuals with or at risk of knee osteoarthritis found that pain pressure threshold trajectories were relatively stable over time, and groups with a low pain pressure threshold (suggesting the presence of nociplastic pain) had worse pain and function at nine years' follow-up.[3] The assessment of nociplastic pain does not occur routinely in clinical settings; thus, another study sought to address this gap by identifying low-burden methods for its assessment in rheumatology practice.[4] Two measures (tender swollen joint count difference and proportion of subjective components over the total Disease Activity Score in 28 joints) were modestly associated with quantitative sensory testing measures of nociplastic pain. These results suggest that discordance between patient-reported and physician-assessed measures of disease activity may reflect an element of nociplastic pain.

Another theme focused on the biopsychosocial nature of pain in rheumatic disease. Cox and colleagues contributed a thorough review of the longitudinal relationship between pain and depression in IA.[5] Key findings were that depression is common among individuals with IA (15%–39%), and as has been commonly found in noninflammatory conditions, there appears to be a bidirectional relationship between pain and depression. Other studies in this issue highlight the relationship of social determinants of health with pain severity, the effectiveness of acceptance and commitment therapy for pain reduction (delivered via a digital health application) on pain in axial spondyloarthritis (axSpA), and the use of natural language processing methods to characterize the psychological, social, and cultural facets of pain discussed by individuals with lupus in an online forum.[6–8]

A third theme centered on opioid use in rheumatic diseases. In a study of individuals with idiopathic inflammatory myopathy, about one in four were taking opioids.[9] Two other studies focused on opioid use among individuals with axSpA. One study found that over approximately the last 20 years, opioid prescriptions remained stable over time in the United Kingdom and decreased slightly in the United States.[10] The other study found that early use of tumor necrosis factor inhibitors was not associated with reduced risk of long-term opioid use.[11] Overall, studies highlight that there is still a high prevalence of opioid use among individuals with rheumatic diseases. This is particularly problematic when opioids are prescribed to address pain that persists in well-

[1]Department of Anesthesiology, University of Michigan, Ann Arbor; [2]Division of Rheumatology, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia; [3]Department of Medicine and Thurston Arthritis Research Center, University of North Carolina at Chapel Hill; [4]Center of Innovation to Accelerate Discovery and Practice Transformation, Department of Veterans Affairs Health Care System, Durham, North Carolina.

Address correspondence via email to Afton L. Hassett, PsyD, at afton@med.umich.edu.

Submitted for publication December 1, 2025; accepted December 2, 2025.

managed IA. Evidence supporting the use of opioids for nociplastic pain is limited, and available evidence suggests opioids are generally not effective for centrally mediated pain and may worsen pain sensitivity in some cases.

Finally, studies in this issue addressed various methodologic issues around pain measurement in both research and clinical settings. A review by Chen and colleagues described important methodologic issues of regression to the mean and contextual effects in clinical trials for pain in rheumatic diseases.[12] This review provides study design and analysis considerations that researchers can implement to better mitigate and account for these challenges. One study in this issue sought to develop a tool to predict progression to difficult-to-treat rheumatoid arthritis.[13] Patients with worse functioning, pain, fatigue, and disease activity had an increased risk of progressing to difficult-to-treat rheumatoid arthritis. However, the tool demonstrated only moderate predictive ability, indicating a need for additional research to optimally predict progression.

These themes echo those from the pain field more broadly, which has been experiencing increased investment at the federal level, resulting in a vast expansion of projects and investigators. This so-called golden age of pain research is largely the direct product of the opioid overdose epidemic. As such, research priorities have focused on developing safe and effective nonopioid treatments, broadening our understanding of pain mechanisms and mechanistically informed interventions, the benefits of a whole-person approach to pain assessment and management, and improving pain research methodology. Overall, the management of pain in rheumatologic settings has evolved in the last two decades in light of new research and will continue to do so as findings from the National Institutes of Health Help End Addiction Long-Term (HEAL) Initiative and many other pain research projects are made public. Together, the articles in this issue underscore the need for integrated approaches that address biologic, psychological, and social contributors to pain to improve outcomes across rheumatic diseases.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Kelleher EM, Meouchi R, Irani A. Beyond inflammation: why understanding the brain matters in inflammatory arthritis. Arthritis Care Res (Hoboken) 2026;78(1):3–14.
2. Rutter-Locher Z, Norton S, Taylor JL, et al. Assessment of pain types in recently diagnosed patients with inflammatory arthritis. Arthritis Care Res (Hoboken) 2026;78(1):154–163.
3. Aoyagi K, Wang N, LaValley M, et al. Longitudinal trajectories of pain sensitization over nine years in people with or at risk of knee osteoarthritis. Arthritis Care Res (Hoboken) 2026;78(1):113–122.
4. Lin CP, Aydemir B, Song J, et al. Association between discordance of disease activity indices and quantitative sensory testing measures of nociplastic pain in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2026;78(1):47–53.
5. Cox N, Farooq S, Twohig H, et al. Longitudinal relationship between pain and depression in people with inflammatory arthritis: a narrative review. Arthritis Care Res (Hoboken) 2026;78(1):15–30.
6. Calvo Pascual LA, Corredor DC, Collantes-Estevez E, et al. Impact of social factors on visual analog scale pain in patients with spondyloarthritis: a propensity score matching analysis. Arthritis Care Res (Hoboken) 2026;78(1):96–105.
7. Kiefer D, Sonkaya Y, Kraus D, et al. Digital behavioral therapy improves outcome in patients with axial spondyloarthritis and persistent pain: a randomized controlled trial. Arthritis Care Res (Hoboken) 2026;78(1):85–95.
8. Walker A, Leung J, Alagappan A, et al. Centering patient voices in lupus pain: a biopsychosocial analysis of Reddit narratives using large language models. Arthritis Care Res (Hoboken) 2026;78(1):123–133.
9. Saygin D, Malfait AM, Ritz EM, et al. Prevalence of pain and factors associated with pain in patients with idiopathic inflammatory myopathies. Arthritis Care Res (Hoboken) 2026;78(1):134–140.
10. Degirmenci HB, Peloquin CE, Westerland M, et al. Trends in opioid prescriptions in individuals with axial spondyloarthritis in the United States and United Kingdom in the last two decades: analysis of electronic health records and insurance claims data. Arthritis Care Res (Hoboken) 2026;78(1):66–75.
11. Petrow E, Meade-Aguilar JA, Peloquin C, et al. Effect of early tumor necrosis factor inhibitor initiation on chronic opioid use in individuals with axial spondyloarthritis. Arthritis Care Res (Hoboken) 2026;78(1):76–84.
12. Chen YT, Zhu G, Hassett AL, et al. Considerations for issues of regression to the mean and contextual effects in clinical trials for pain in rheumatic diseases. Arthritis Care Res (Hoboken) 2026;78(1):31–37.
13. Paudel ML, Shadick N, Weinblatt M, et al. Development and external validation of a multivariable predictive model for progression to difficult-to-treat rheumatoid arthritis in biologic-experienced patients. Arthritis Care Res (Hoboken) 2026;78(1):54–65.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 3–14
DOI 10.1002/acr.25694

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

Beyond Inflammation: Why Understanding the Brain Matters in Inflammatory Arthritis

Eoin M. Kelleher,[1,2] Rosario Meouchi,[3] and Anushka Irani[1,3]

Persistent pain remains a major challenge in inflammatory arthritis, even when joint inflammation is well controlled. Pain and associated symptoms such as fatigue cannot be explained by peripheral inflammation alone but reflect altered central pain processing. These changes may arise through "top-down" mechanisms, reflecting pre-existing dysfunction in pain perception, or "bottom-up" pathways, driven by peripheral inflammation acting on the brain. Neuroimaging has transformed understanding of these processes by providing in vivo markers of how brain function and structure are related to pain. Functional magnetic resonance imaging (MRI) demonstrates that both task-evoked and resting-state activity are altered in inflammatory arthritis. Connectivity changes involving the thalamus, insula, medial prefrontal cortex, and default mode and salience networks correlate with pain, fatigue, and affective symptoms. Notably, tumor necrosis factor α (TNF-α) inhibitors rapidly normalize pain-related activation, preceding improvements in joint swelling, strongly supporting a bottom-up role for peripheral inflammation. Recent randomized controlled trial data show that baseline central nervous system pain activation predicts analgesic response to TNF-α blockade, positioning neuroimaging as a potential tool for treatment stratification. Complementary modalities provide further insights. Proton electron tomography studies suggest altered pain responses, and novel tracers may clarify contributions of neuroinflammation. Magnetic resonance spectroscopy reveals neurochemical correlates such as increased choline and myo-inositol linked to fatigue, although group-level evidence for overt neuroinflammation remains limited. Structural MRI highlights gray matter changes in regions mediating sensory, cognitive, and affective processing. Together, this supports a dual top-down and bottom-up model of persistent pain in inflammatory arthritis, with important implications for mechanism-based therapies targeting both immune and brain pathways.

Introduction

Pain is the cardinal symptom of inflammatory arthritis and remains a top priority for patients living with conditions such as rheumatoid arthritis (RA), psoriatic arthritis (PsA), and axial spondyloarthritis (axSpA).[1,2] Despite significant advances in pharmacologic therapies targeting immune pathways, many patients continue to experience persistent pain even when inflammatory joint disease is apparently well controlled.[3]

Although inflammation contributes to pain intensity, improved control of synovitis has not consistently translated into meaningful reductions in pain or functional limitation over time.[4,5] This clinical disconnect highlights a critical unmet need and calls into question long-standing assumptions that ongoing pain necessarily directly reflects active joint inflammation.

The emergence of the "difficult-to-treat" RA concept underscores this challenge,[6–8] estimated to affect 14.4% of patients with RA and up to 22.3% of those treated with biologics.[9] Patients may present with persistent symptoms suggestive of disease activity despite adequate immunomodulatory therapy with multiple conventional and biologic disease-modifying antirheumatic drugs (DMARDs). Importantly, recent efforts to subcategorize these patients into **inflammatory-refractory RA** and **noninflammatory RA**[6] recognize the possibility of distinct underlying pain mechanisms.[10,11]

This mechanistic reframing is mirrored in the updated chronic pain taxonomy introduced by the International Association for the Study of Pain (IASP) and adopted into the *International Classification of Diseases, 11th Revision*.[12] Chronic pain is now classified as either chronic primary (eg, fibromyalgia) or chronic secondary (eg, pain because of inflammatory arthritis), with the latter designation capturing cases in which pain persists independently of active inflammation. This classification has the potential to inform

[1]University of Oxford, Oxford, UK; [2]Massachusetts General Hospital, Boston; [3]Mayo Clinic Florida, Jacksonville.

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25694.

Address correspondence via email to Eoin M. Kelleher, MBBCh, MSc, DPhil, at ekelleher11@mgh.harvard.edu.

Submitted for publication August 7, 2025; accepted in revised form October 13, 2025.

more tailored treatment strategies for patients with inflammatory arthritis and coexisting chronic pain.

Importantly, multiple pain mechanisms can coexist in individuals with inflammatory arthritis. The IASP also proposes three mechanistic descriptors: nociceptive, neuropathic, and nociplastic pain.[13,14] Nociceptive pain arises from tissue damage and is typical in active synovitis. Neuropathic pain, caused by somatosensory system lesions, is relatively uncommon in these conditions, although features of neuropathic pain may be present.[15] Nociplastic pain reflects altered central pain processing without clear evidence of ongoing tissue damage or nerve injury and may underlie persistent pain in patients with well-controlled inflammation. This category overlaps conceptually with previously used terms, such as "central sensitization" or "centralized pain," with fibromyalgia serving as a prototypical example.[11,13] Recent large-scale neuroimaging data further support the biologic validity of nociplastic pain. In a population-based study of >40,000 participants in UK Biobank, we demonstrated that higher Fibromyalgia Index scores, reflecting greater nociplastic pain burden, were associated with altered structural connectivity among the periaqueductal gray, amygdala, and hypothalamus, key nodes of the descending pain modulatory system (DPMS).[16] These findings provide population-level evidence for central pain amplification mechanisms.

There is accumulating evidence to suggest that persistent pain in inflammatory arthritis may have a nociplastic, in addition to nociceptive, component. Proinflammatory cytokines may influence neural activity at both peripheral and central levels.[17,18] Neuroimaging offers a promising avenue to explore these phenomena in vivo, allowing noninvasive assessment of brain mechanisms underlying pain and other centrally mediated symptoms.[19–21]

This review aims to summarize key neuroimaging findings related to inflammatory arthritis pain, examine how current treatment strategies influence central pain processing, and propose a more integrated, mechanism-based clinical model to guide care. By recognizing the brain as central to inflammatory arthritis pathophysiology, we advocate a shift from inflammation-focused "treat-to-target" strategies toward a broader goal of "treat-to-recovery," emphasizing restoration of function and quality of life.

Imaging the pain experience: a window into the brain

Advances in neuroimaging have provided powerful tools to investigate the neural mechanisms underlying pain in inflammatory arthritis. These tools offer complementary perspectives on how the brain processes pain and related symptoms such as fatigue, mood disturbance, and cognitive dysfunction, creating a multidimensional picture of the pain experience.

Functional magnetic resonance imaging. Blood oxygenation level–dependent (BOLD) functional magnetic resonance imaging (fMRI) is the most widely used method to study pain-related brain activity. Task-based fMRI can reveal how the brain responds to experimentally evoked pain, often showing increased activation in key pain-processing regions such as the insula, anterior cingulate cortex (ACC), and thalamus. Resting-state fMRI (rs-fMRI), by contrast, measures spontaneous connectivity among brain regions, allowing assessment of how large-scale networks such as the default mode network (DMN), salience network, and frontoparietal network interact in chronic pain states.

Positron emission tomography. Positron emission tomography (PET) provides complementary information on brain function and, with specific radiotracers, can assess both regional blood flow and markers of neuroinflammation. In chronic pain conditions, PET has highlighted altered nociceptive processing and, with newer tracers, evidence of glial activation, suggesting a contribution of neuroimmune mechanisms to persistent pain.[22]

Magnetic resonance spectroscopy. Magnetic resonance spectroscopy (MRS) enables in vivo assessment of brain neurochemistry by quantifying metabolites such as choline, myo-inositol, glutamate, and γ-aminobutyric acid. These measures can provide insights into cellular processes including glial activity and neurotransmission, which are thought to play a role in sustaining chronic pain.

Structural MRI. High-resolution structural MRI allows volumetric analysis of gray and white matter, offering insight into long-term brain changes. Alterations in regions involved in pain appraisal, affective processing, and sensorimotor integration have been reported across chronic pain populations.

Together, these modalities provide a multidimensional view of the pain experience in inflammatory arthritis, from real-time network activity and connectivity to underlying neurochemical and structural changes. By integrating these approaches, we can begin to unravel the complex interactions among inflammation, nociception, and central pain modulation.

Pain pathways rewired: neuroimaging insights into central mechanisms in inflammatory arthritis

Neuroimaging studies provide compelling evidence that persistent pain in inflammatory arthritis is not just caused by peripheral inflammatory mechanisms but also involves central nervous system (CNS) pathways in its maintenance. These findings are particularly relevant for the subset of patients whose pain persists despite optimal control of joint inflammation.

Evidence for nociplastic pain and overlap with fibromyalgia. fMRI studies indicate that RA can represent a

"mixed pain state," in which nociplastic pain mechanisms in the CNS coexist alongside nociceptive drivers of pain from the joints (Figure 1). Basu and colleagues demonstrated that patients with RA with higher "fibromyalgianess" scores, even without meeting the diagnostic criteria for fibromyalgia, exhibited abnormal functional connectivity between the DMN and the insula, patterns previously identified as characteristic of fibromyalgia.[23] Notably, these findings were associated not only with widespread pain but also with fatigue and cognitive dysfunction, underscoring the multisystem symptom burden often observed in these patients. These findings align with quantitative sensory testing data showing lower pain pressure thresholds across both affected and unaffected body regions in RA, consistent with central sensitization.[24,25] In a recent study in PsA, Sunzini and colleagues reported similar results, finding that fibromyalgianess scores correlated positively with resting-state functional connectivity found between the right anterior insula and the DMN in rs-fMRI.[26] However, in contrast, studies by Flodin and colleagues and Sandström and colleagues found that CNS activation to evoked pain at nonjoint sites did not differ between those with RA and healthy controls.[27,28] This suggests that although pain sensitivity is increased in joint sites, there is not necessarily a generalized increase in peripheral pain sensitivity. However, it should be noted that these studies recruited patients with RA without comorbid fibromyalgia symptoms and thus may have excluded those patients with RA with more severe nociplastic features.

Suggestion of neuropathic pain characteristics. Wu and colleagues investigated the presence of a neuropathic pain component in patients with ankylosing spondylitis (AS).[29] Using

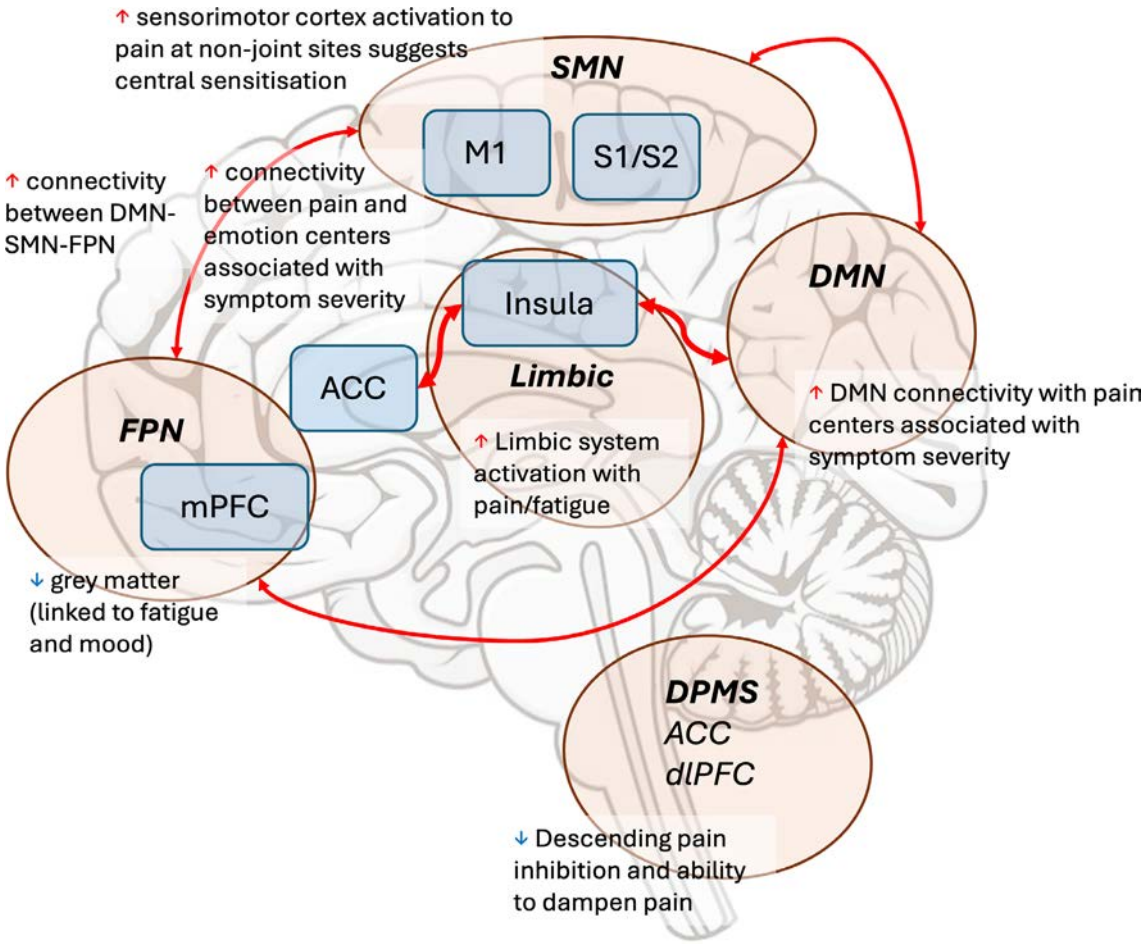


Figure 1. Brain mechanisms of nociplastic pain in inflammatory arthritis. Neuroimaging studies show that patients with inflammatory arthritis may experience persistent pain even when inflammation is controlled because of changes in how the brain processes pain. Brain networks that normally regulate sensation (SMN), thinking (FPN), and self-awareness (DMN) become more strongly interconnected, contributing to fatigue and pain. Connections between pain and emotion centers (insula and ACC) are linked to symptom severity, whereas the limbic system shows greater activation with pain and fatigue. Exaggerated pain responses extend beyond inflamed joints, suggesting central sensitization. At the same time, brain regions that normally dampen pain (DPMS) show reduced activity. Shrinkage of gray matter in the mPFC further relates to fatigue and mood changes. These findings highlight brain contributions to nociplastic pain in inflammatory arthritis. ACC, anterior cingulate cortex; dlPFC, dorsolateral prefrontal cortex; DMN, default mode network; DPMS, descending pain modulatory system; FPN, frontoparietal network; mPFC, medial prefrontal cortex; SMN, sensorimotor network.

structural MRI and the painDETECT questionnaire (a screening tool for neuropathic pain that was validated in patients with low back pain[30]) they found an inverse correlation between left primary somatosensory cortex thickness and painDETECT scores. Conversely, thickness of the left primary motor cortex, right ACC, and right dorsolateral prefrontal cortex (dlPFC) exhibited a direct correlation with painDETECT scores. Although definite neuropathic pain cannot be identified using the painDETECT alone, as this requires the confirmation of a somatosensory lesion,[31] the association between these neuropathic-like pain characteristics and changes in centrally mediated pain processing has been demonstrated by us and others in a variety of musculoskeletal conditions.[32–35]

Altered resting-state connectivity and pain-processing pathways. Resting-state connectivity analyses further support the presence of widespread CNS reorganization in inflammatory arthritis. Flodin and colleagues observed increased PFC-insula connectivity at rest in those with RA compared with healthy controls, however this was not itself associated with symptom severity, which may be caused by the relative insensitivity of rs-fMRI in small sample sizes.[27] Schrepf and colleagues reported that higher levels of systemic inflammation in RA were associated with enhanced connectivity among the medial PFC (mPFC), inferior parietal lobule (IPL), and canonical brain networks, including the DMN and dorsal attention network, during an attention task.[36] These changes correlated with greater fatigue, pain, and cognitive dysfunction, suggesting that inflammation may act as a driver of maladaptive network changes in the CNS.[36] Graph theoretical analyses extended these findings, revealing a distributed network of 49 regions linked to peripheral inflammation, measured by erythrocyte sedimentation rate (ESR), further implicating the IPL and mPFC as important nodes in the translation of inflammatory signals into central pain-processing abnormalities.

These observations complement longitudinal imaging work demonstrating that anti–tumor necrosis factor (TNF) therapy reduces pronociceptive brain connectivity, in some cases preceding improvements in joint swelling and clinical indices of disease activity.[37,38] This is supported by the recent PreCePRA randomized multicenter clinical trial by Hess and colleagues,[39] which investigated whether baseline brain activation on task-based fMRI could predict treatment response to certolizumab pegol. Patients with RA were stratified by the extent of brain activation on fMRI in response to joint compression and randomized to anti-TNF therapy or placebo. At 12 weeks, more than half of patients with high baseline CNS activation achieved low disease activity on TNF inhibition compared with less than half in the low-activation group and only one-quarter of those receiving the placebo. Importantly, the improved outcomes in the high-activation group were largely driven by patient-reported measures such as pain and global disease activity, rather than objective markers of inflammation, which improved similarly across treatment groups. Functional imaging further demonstrated that TNF blockade selectively reduced pain-related activation in the primary somatosensory cortex, motor cortex, and frontal pole in the high-activation group. These findings suggest that a strong "disease twin" in the CNS, reflecting heightened central representation of joint pain, may identify patients who are more responsive to TNF inhibition, supporting the concept that bottom-up inflammatory signaling amplifies central sensitization. Conversely, patients with low CNS activation may have pain sustained more by top-down processes and respond less to purely anti-inflammatory strategies. These findings imply that TNF-α may influence pain through mechanisms partially independent of its effects on synovial inflammation, with one pathway driving joint swelling and another promoting peripheral sensitization and central amplification of pain. Collectively, the previous observational work and this recent trial provide proof of concept that neuroimaging biomarkers could stratify patients with inflammatory arthritis for mechanism-based treatment approaches.

Together, these neuroimaging insights illustrate that persistent pain in inflammatory arthritis cannot be solely explained by peripheral inflammation. Instead, aberrant connectivity across salience network, DMN, and frontoparietal network, as well as structural changes in regions such as the mPFC and IPL, highlight the CNS's active role in shaping the pain experience. Recognizing these mechanisms is important for rheumatologists, as they provide a rationale for integrating interventions that target pain processing in the CNS alongside traditional immunosuppressive strategies.

When inflammation talks to the brain: the interaction between inflammation and central pain pathways

Growing evidence suggests that peripheral inflammation in RA and related conditions communicates directly with the CNS, shaping both functional connectivity and neurochemical signatures (Figure 2). The resulting brain changes help explain why patients frequently experience fatigue, mood disturbance, and pain even in the absence of overt synovitis.

Neuroinflammation and immune-to-brain signaling. Peripheral inflammatory mediators, including TNF-α, interleukin (IL)-1, and IL-6, can signal to the CNS via multiple pathways: humoral transport across the blood-brain barrier, activation of afferent vagal fibers, and stimulation of circumventricular organs. These signals may promote microglial activation and subsequent release of proinflammatory cytokines within the brain; however, evidence for the presence of neural inflammation is lacking.[40,41] Chronic activation of these pathways contributes to the cluster of "sickness behaviors" (fatigue, anhedonia, and cognitive dysfunction) that typify many inflammatory rheumatic diseases.[42]

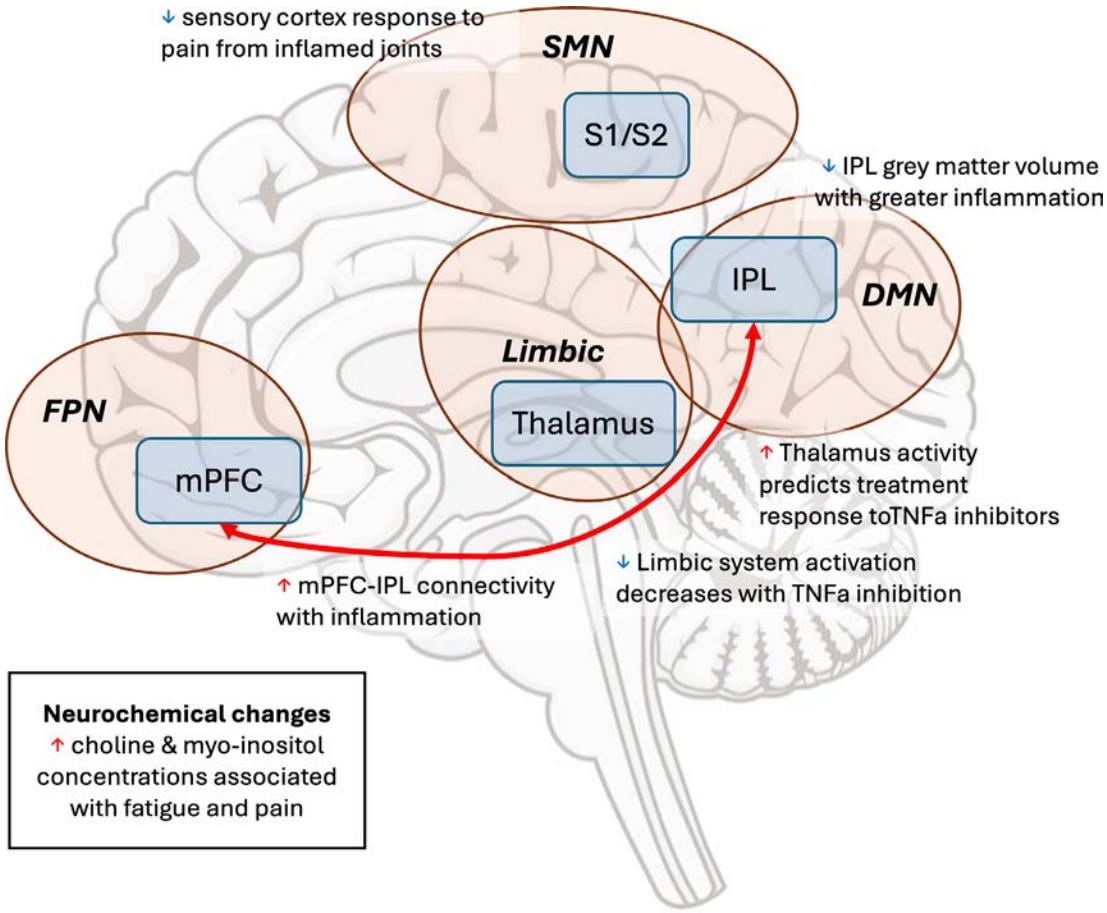


Figure 2. Brain changes linked to active inflammation in inflammatory arthritis. Neuroimaging studies show that active joint inflammation affects pain processing in the brain. Reduced activity in the sensory cortex (S1/S2) is seen when pain is evoked at inflamed joints. Activity in the thalamus, a key relay in pain perception, may predict response to TNF-α inhibitors. Limbic system activity, associated with pain and fatigue, decreases after TNF blockade, even before joint swelling improves. Increased connectivity between the mPFC and IPL, a key node of the DMN, is linked to systemic inflammation. Neurochemical changes, such as higher choline and myo-inositol, are associated with pain and fatigue, and loss of gray matter in the IPL correlates with greater inflammation. These findings demonstrate that inflammation drives both functional and structural brain changes, some of which may improve with treatment. DMN, default mode network; FPN, frontoparietal network; IPL, inferior parietal lobule; mPFC, medial prefrontal cortex; SMN, sensorimotor network; TNF-α, tumor necrosis factor α. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25694/abstract.

Although the concept of sickness behavior usefully encapsulates the constellation of neurovegetative and affective symptoms that accompany immune activation, accumulating evidence indicates that the physiologic and behavioral responses to inflammation evolve over time and differ between the acute and chronic phases. Work by Capuron and colleagues, among others, of patients undergoing chronic interferon-α (IFN-α) therapy have been particularly informative in this respect.[43] During early treatment, patients frequently experience fatigue, psychomotor slowing, anorexia, and sleep disturbance, features that align closely with classic sickness behavior, whereas mood and cognitive symptoms such as depression, anxiety, and attentional deficits emerge more gradually with continued cytokine exposure.[44–48] Neuroimaging work demonstrates that these phases map onto distinct neurobiological substrates: acute IFN-α administration is associated with hyper-reactivity of the hypothalamic-pituitary-adrenal axis, particularly in individuals who later develop major depression, whereas sustained treatment induces altered activity in fronto-subcortical circuits.[49] PET and fMRI studies show increased basal ganglia glucose metabolism and reduced ventral striatal responsiveness to reward, consistent with dopaminergic dysregulation underlying fatigue and anhedonia.[50,51] Concurrently, heightened activation of the ACC during cognitive tasks suggests compensatory increases in effortful control under inflammatory load.[49] Together, these findings indicate that chronic cytokine exposure drives a progressive reorganization of central stress-response and reward networks, distinguishing later-phase inflammatory psychopathology from the transient, adaptive sickness response typical of acute inflammation.

Peripheral inflammation linked to altered connectivity. fMRI studies demonstrate that systemic

inflammation is associated with large-scale alterations in resting-state connectivity. In RA, higher ESR and C-reactive protein (CRP) levels were linked to stronger connectivity among the mPFC, IPL, and multiple networks including the DMN, salience network, and dorsal attention network. These connectivity changes were associated with levels of fatigue, pain severity, and cognitive impairment, with the IPL appearing as a key hub in this "inflammation configuration." Reduced gray matter volume (GMV) in the IPL was also associated with higher inflammation, indicating that systemic immune activity contributes not only to functional reorganization but also to structural changes.[36]

Beyond group-level associations between systemic inflammation and neural activity, there is increasing recognition that the neuroimmune interface exhibits substantial interindividual variability. Even among patients with similar degrees of peripheral inflammation, central consequences may differ markedly depending on vulnerability factors within neural and immune networks. For instance, Kaplan and colleagues reported that in RA, ESR was associated with pronociceptive functional connectivity only in individuals exhibiting fibromyalgia-like symptoms, suggesting that susceptibility to the central effects of inflammation varies among patients.[52] Such variability may help explain the heterogeneity of pain, fatigue, and mood symptoms in rheumatic disease and supports the need for stratified, mechanism-based approaches to therapy.

Neurochemical correlates of systemic inflammation. MRS has provided complementary insights. Although no widespread metabolic abnormalities were identified in patients with RA compared with healthy controls, brain choline and myo-inositol, markers of cell turnover and glial proliferation, respectively, were positively correlated with fatigue severity and with joint swelling and tenderness.[40] These findings suggest that even in the absence of overt neuroinflammation on group comparisons, neurochemical alterations in subsets of patients may contribute to central symptoms.

Impact of cytokine blockade on central pain processing. Neuroimaging studies have shown that anti-TNF therapy produces rapid and measurable effects on CNS pain processing. For example, blockade of TNF-α leads to reductions in the activation of limbic structures, the thalamus, and somatosensory cortex within 24 hours, often preceding measurable improvements in joint swelling.[38] A recent preprint has shown similar findings, with baseline thalamic activation predicting clinical response, raising the possibility that neuroimaging could be used as a biomarker for treatment stratification.[53] Similarly, Abe and colleagues demonstrated that aberrant functional connectivity between the ACC and insula was associated with treatment response across both RA and SpA, highlighting shared central mechanisms in inflammatory arthritis.[54]

In addition, TNF inhibition has been associated with structural brain changes in other rheumatic conditions. Wu and colleagues demonstrated a correlation between pain reduction in patients with AS and alterations in cortical thickness.[55] Specifically, greater pain reduction was associated with cortical thickening in the left orbitofrontal cortex and thinning in the right somatosensory cortex, as measured by structural MRI. These cortical changes may relate to the emotional and sensory aspects of pain, respectively, with orbitofrontal thickening linked to emotional pain processing and somatosensory thinning associated with central sensitization.[56,57]

Building on the evidence for the central effects of TNF blockade, JAK inhibitors have also demonstrated striking analgesic benefits in RA, further highlighting the potential for immune-targeted therapies to modulate pain beyond their anti-inflammatory effects.[58] Clinical trials of tofacitinib, baricitinib, and upadacitinib consistently show rapid reductions in pain, in some cases greater than with TNF inhibitors.[59] Importantly, post hoc analyses indicate that these improvements are only partly explained by reductions in swollen joints or markers of inflammation. For example, in the RA-BEAM trial, baricitinib produced greater pain relief than adalimumab despite comparable control of synovitis, suggesting additional mechanisms influencing central pain pathways.[60] Similarly, Dougados and colleagues reported that tofacitinib reduced residual pain, even in patients who had already achieved low disease activity or remission, supporting analgesic effects independent of overt inflammation.[61] These clinical findings are consistent with preclinical evidence that proinflammatory cytokines such as TNF-α, IL-6, and IL-17 can directly sensitize nociceptive neurons and promote both peripheral and central sensitization, thereby sustaining hyperalgesia independent of joint swelling.[18] Although direct neuroimaging data on JAK inhibition remain lacking, these clinical observations parallel the CNS changes seen with TNF blockade and raise the possibility that JAK inhibitors also attenuate pronociceptive central signaling. By targeting multiple cytokine pathways converging on JAK/STAT signaling, these agents may influence both bottom-up pathways driven by peripheral inflammation and top-down mechanisms of central sensitization. Future neuroimaging studies are needed to determine whether JAK inhibitors normalize pain-related network activity in a manner similar to TNF-α blockade.

Clinical implications. Together, these findings demonstrate that systemic inflammation shapes brain circuits underlying pain, fatigue, and cognition, and that modulation of these circuits may underlie part of the benefit of biologic therapy. For clinicians, this explains why patients may report rapid symptomatic improvement following cytokine blockade, sometimes even before peripheral measures of inflammation have fully resolved. They also suggest that persistent CNS changes may perpetuate symptoms despite good control of synovitis, reinforcing the need

for mechanism-based approaches to pain management in inflammatory arthritis.

What happens in the brain when we treat the joints?

Although most pharmacologic therapies for inflammatory arthritis target immune pathways within the joints, neuroimaging studies demonstrate that their benefits can also extend to the CNS. A small but growing body of research has examined how biologic DMARDs, particularly TNF-α inhibitors, modulate brain circuits involved in pain perception.

Rapid central effects of TNF-α blockade. Early fMRI work highlighted that TNF-α blockade can rapidly normalize pain-related brain activity, often before clinical measures of joint inflammation improve. Hess and colleagues observed a significant reduction in activity in the thalamus, insula, and ACC within 24 hours of infliximab infusion in patients with RA with active disease.[38] Similarly, Rech and colleagues reported that responders to certolizumab pegol demonstrated pronounced reductions in the activation of the somatosensory cortex, dlPFC, and limbic structures as early as 3 days after treatment.[37] These early CNS changes occurred well before reductions in Disease Activity Scores in 28 joints scores were evident, suggesting that the CNS is highly responsive to cytokine blockade.

Differentiating responders and nonresponders. Neuroimaging has also provided insight into why some patients respond to TNF-α inhibitors and others do not. Rech and colleagues showed that TNF-α inhibitor responders exhibited greater baseline pain-related CNS activation compared with nonresponders.[37] After treatment, responders showed progressive declines in activation across the pain matrix, whereas nonresponders had persistently elevated activity, particularly in the thalamic and periaqueductal gray regions associated with descending modulation. Functional connectivity analyses further revealed that responders had more adaptive changes in thalamic-insula and thalamic-ACC connectivity, whereas nonresponders maintained maladaptive patterns.

In a similar vein, Espartal and colleagues conducted a small proof-of-concept study to assess the use of fMRI to evaluate pain in response to TNF-α inhibitors in six patients with PsA.[62] In this study, patients exhibiting a poor treatment response showed baseline activation of the sensory motor cortex. In contrast, patients with a strong treatment response did not have any baseline activation and showed activation in the sensory motor cortex, amygdala, insula, frontal, and prefrontal areas during the following week of treatment. Patients with poor treatment response showed minimal change, but larger studies are needed to verify these findings.

One of the earliest studies to demonstrate the utility of fMRI in inflammatory arthritis was a single-patient pharmacologic-MRI study investigating the modulation of brain activity of PsA pain by a single dose of a COX-2 inhibitor.[63] In this study, the patient was a 53-year-old man with active PsA where it had been noted that treatment with a COX-2 inhibitor had resulted in profound pain and rapid relief without resolution of the associated joint swelling. Brain imaging showed that activity in the anterior and posterior insula correlated with pain intensity, showing reduced activity following COX-2 inhibitor administration. This study, albeit in a single patient, underscores the potential of fMRI as a tool for identifying central mechanisms of pain and its value in advancing mechanism-based treatment stratification.[64]

Connectivity predictors of treatment response. More recent studies suggest that resting-state connectivity patterns may predict clinical outcomes. Abe and colleagues[54] demonstrated that altered insula-ACC connectivity at baseline was associated with subsequent treatment response in both RA and SpA, highlighting the role of salience network dysfunction across inflammatory joint conditions. In a prospective cohort, Wartolowska and colleagues[53] reported that higher baseline thalamic activation predicted greater improvement in pain following TNF blockade, further supporting the potential utility of neuroimaging biomarkers in stratifying patients for therapy.

Clinical implications. These findings provide a neurobiologic explanation for the frequently observed phenomenon in clinical practice that some patients report rapid pain relief shortly after initiation of TNF inhibitors, even in the absence of objective improvements in joint swelling or CRP. They also underscore that persistent pain after effective immunosuppression may reflect enduring central mechanisms rather than refractory synovitis. In the future, neuroimaging markers may help to personalize treatment decisions by identifying patients most likely to benefit from cytokine blockade and guiding the integration of adjunctive therapies targeting nociplastic pain pathways.

Beyond pain: central mechanisms behind fatigue and mood disturbance

Although pain is the most prominent and burdensome symptom in inflammatory arthritis, patients frequently report fatigue, low mood, and reduced resilience as equally disabling.

Depression and altered affective circuits. Neuroimaging work underscores the link between depressive symptoms and altered pain processing in inflammatory arthritis. Schweinhardt and colleagues found that in RA, activation of the mPFC during evoked joint pain correlated with greater depressive symptom scores, suggesting that affective state directly influences cerebral pain processing.[65] Connectivity among the

mPFC, posterior cingulate cortex, and other limbic structures was strengthened in patients with higher depressive symptoms, providing a neural correlate for the clinical observation that depression amplifies pain experience. Kaplan and colleagues similarly emphasized the role of limbic hyperactivation in the affective dimension of pain, reinforcing the concept that depression and pain share overlapping neurobiologic substrates.[52]

Additionally, rs-fMRI studies can provide valuable insights into the neurobiologic mechanisms underlying the relationship between brain activity and pain and depression. Hua and colleagues investigated regional homogeneity (ReHo), a measure of functional connectivity alterations in patients with AS.[66] Compared with healthy controls, their study revealed higher ReHo in the left precuneus and right middle frontal gyrus and reduced ReHo in the left superior temporal gyrus and right paracentral lobule. A negative correlation was observed between ReHo in the right paracentral lobule and disease duration and back pain scores. Conversely, a positive correlation was observed between ReHo in the left precuneus and depression and back pain scores.

Fatigue and network reorganization. Fatigue is among the most disabling symptoms reported by patients with inflammatory arthritis and often persists despite good inflammatory control. Schrepf and colleagues demonstrated that systemic markers of inflammation (ESR and CRP) were associated with altered connectivity among the mPFC, IPL, and large-scale networks such as the DMN and dorsal attention network.[36] These connectivity changes correlated not only with pain but also with fatigue and subjective cognitive dysfunction, suggesting a shared neurobiologic basis for these symptoms. In addition, Liu and colleagues reported similar findings.[67] In a study that investigated functional and structural alterations in patients with AS, they observed that altered functional connectivity between DMN and salience network nodes correlated with fatigue. For within DMN nodes this was correlated not only with pain but also with fatigue. Furthermore, putamen GMV was associated with pain and fatigue. Hua and colleagues also observed an increase of GMV in this subcortical brain region in patients with AS with moderate back pain; this increase correlated with disease duration.[68]

Sex differences. The experience of pain in inflammatory arthritis is known to differ between sexes,[69–71] but the underlying mechanistic drivers are not yet fully understood. Neuroimaging can further shed light on the distinct pain mechanisms between sexes, which in turn may help to uncover tailored treatment strategies. Fauchon and colleagues used rs-fMRI data to conduct modular analysis using graph theory and found distinct brain modular structures in healthy male and female patients.[72] Data from patients with AS with chronic pain revealed altered brain organization, particularly in female patients, who exhibited DMN modularity, increased salience/attention network activity, and decreased frontal network connectivity. In contrast, male patients exhibited increased intermodular connectivity, particularly between the DMN and prefrontal areas. Similarly, Rogachov and colleagues investigated the relationship between chronic pain in patients with AS and low-frequency oscillations (LFOs) in the BOLD signal of rs-fMRI.[73] LFOs, which are measured at rest, reflect slow, rhythmic fluctuations in the BOLD signal, providing insights into brain network function.[74] Compared with healthy controls, patients with AS exhibited increased LFO amplitudes and distinct LFO patterns within pain-related brain networks. Analysis of sex differences revealed that female patients (except in the posterior cingulate/precuneus) exhibited the slowest brain frequencies in both the healthy control and AS groups. Within the AS group, male and female patients displayed different LFO patterns despite reporting similar pain levels. Overall, these findings suggest that, even if the burden of pain is similar between sexes, targeting the different underlying brain-based mechanisms may improve overall treatment responses.

Resilience and adaptive central responses. Emerging evidence highlights that not all patients respond to systemic inflammation in the same way. Some individuals appear to maintain functional resilience, with preserved or adaptive connectivity in frontoparietal and executive networks despite ongoing inflammation. Abe and colleagues[54] identified that connectivity patterns between the ACC and insula not only predicted treatment response but may also reflect an individual's capacity for central compensation in the face of persistent inflammatory signaling. These findings align with broader literature on cognitive resilience, suggesting that variability in network plasticity and descending pain modulatory capacity may partially explain why some patients adapt better to chronic disease burden than others. Interestingly, psychological resilience has also been linked with pain-related functional connectivity.[75] Hemington and colleagues found that resilience scores of healthy control patients showed a negative correlation with DMN-salience network internetwork connectivity. In contrast, patients with AS had a positive correlation between pain scores and interconnectivity between these networks.

Clinical implications. Together, these studies extend the neuroimaging evidence in inflammatory arthritis beyond pain to encompass fatigue, depression, and resilience. For clinicians, they reinforce that persistent fatigue or mood disturbance should not be dismissed as nonspecific complaints but rather as potential markers of central dysregulation linked to inflammatory signaling. This recognition may encourage integration of mental health support, fatigue-management strategies, and centrally acting pharmacologic or nonpharmacologic interventions into standard care alongside immunosuppressive therapy.

Integrating the pathways: a mechanism-based model for pain management in inflammatory arthritis

The accumulating neuroimaging evidence supports a reframing of pain in inflammatory arthritis as a product of both bottom-up inflammatory signaling and top-down CNS modulation (Figure 3). Although DMARDs remain the cornerstone of treatment for synovitis, many patients continue to experience pain, fatigue, and mood disturbance after inflammatory control is achieved. This disconnect highlights the need for a mechanism-based model of pain management that integrates immune modulation with strategies targeting central pain pathways.

A top-down and bottom-up framework. In this model, peripheral inflammation acts as a bottom-up driver of pain, activating ascending nociceptive pathways via cytokine-mediated effects on the thalamus, insula, and somatosensory cortices.[36,38] At the same time, maladaptive top-down processes, including hyperconnectivity of the DMN, salience network, and limbic network and impaired DPMS activity, sustain nociplastic pain even when inflammation is controlled.[23,37] Importantly, these processes are not mutually exclusive but frequently coexist, producing the complex pain phenotypes observed in RA, PsA, and axSpA.

Opportunities for CNS-targeted interventions. Recognizing the central contribution to pain in inflammatory arthritis opens the door for adjunctive therapies targeting CNS pathways.

- Pharmacologic interventions: centrally acting medications such as serotonin-noradrenaline reuptake inhibitors (eg,

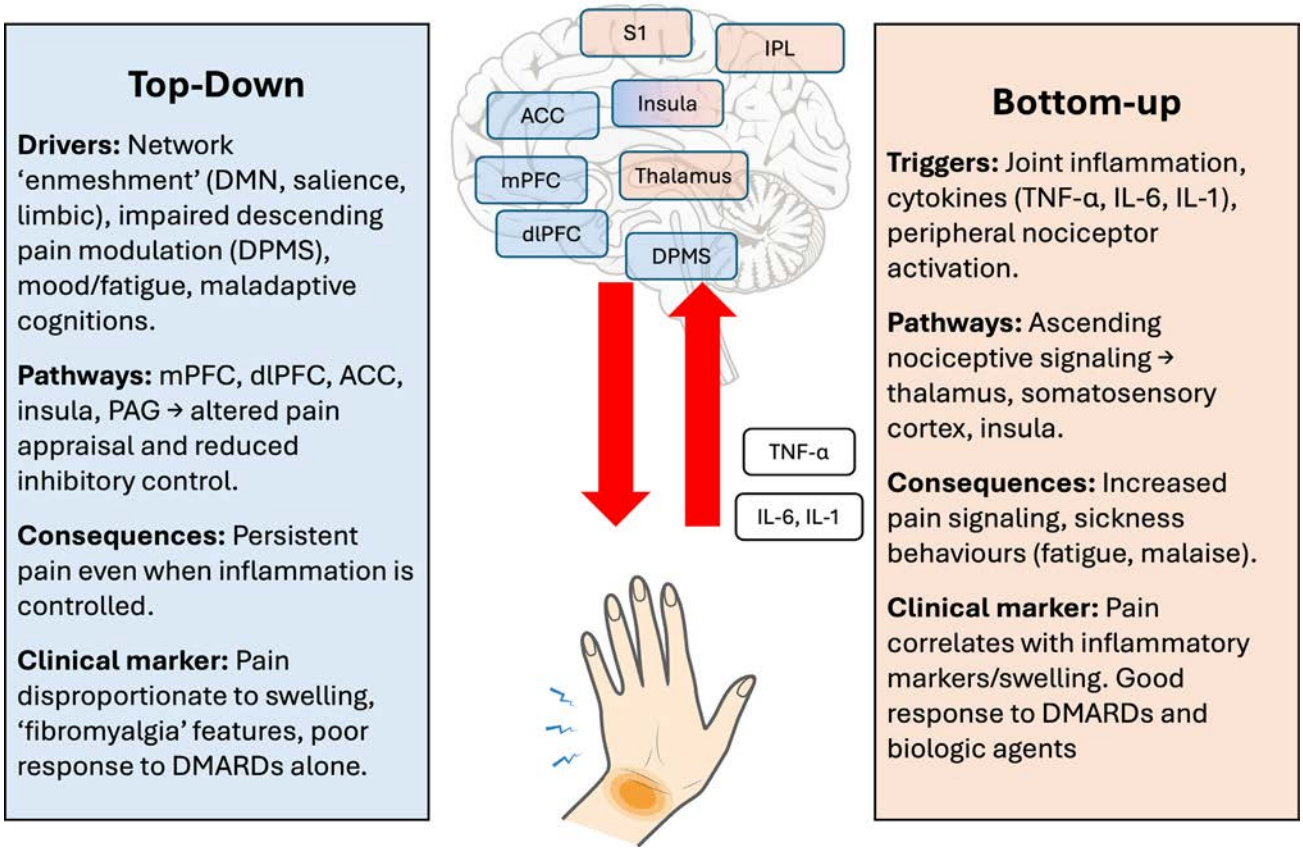


Figure 3. Top-down and bottom-up mechanisms of persistent pain in inflammatory arthritis. Neuroimaging evidence highlights two complementary pathways that contribute to pain beyond joint inflammation. Bottom-up mechanisms involve peripheral drivers such as joint inflammation and cytokines (TNF-α, IL-6, and IL-1) that activate ascending nociceptive pathways to the thalamus, somatosensory cortex, and insula, producing increased pain signaling and sickness behaviors. This phenotype is characterized by pain that correlates with markers of inflammation and typically improves with DMARDs or biologics. In contrast, top-down mechanisms reflect central sensitization and maladaptive modulation of pain, driven by the enmeshment of the default mode, salience, and limbic networks, impaired DPMS activity, and mood/fatigue-related processes. These changes, involving the mPFC, ACC, insula, and PAG, sustain pain despite controlled inflammation. Clinically, this presents as pain disproportionate to swelling, fibromyalgia-like features, and poor response to DMARDs alone. Together, these pathways emphasize the dual central and peripheral origins of persistent pain in inflammatory arthritis. ACC, anterior cingulate cortex; dlPFC, dorsolateral prefrontal cortex; DMARD, disease-modifying antirheumatic drug; DMN, default mode network; DPMS, descending pain modulatory system; IL, interleukin; IPL, inferior parietal lobule; mPFC, medial prefrontal cortex; PAG, periaqueductal gray; TNF-α, tumor necrosis factor α. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25694/abstract.

duloxetine) have demonstrated efficacy in fibromyalgia and may benefit patients with inflammatory arthritis with nociplastic pain features.[52] Other agents, including pregabalin and low-dose tricyclic antidepressants, may also be considered in select cases.

- Psychological therapies: cognitive behavioral therapy and acceptance and commitment therapy have been shown to improve coping, reduce catastrophizing, and enhance resilience, with fMRI studies suggesting they modulate the frontoparietal and limbic circuits implicated in nociplastic pain.[65]
- Lifestyle and nonpharmacologic approaches: exercise, sleep optimization, and stress-reduction strategies (eg, mindfulness-based interventions) have demonstrated benefits for both pain and fatigue. Neuroimaging evidence suggests these interventions enhance functional connectivity within executive and descending inhibitory networks, potentially restoring more adaptive top-down regulation.

Toward mechanism-based personalized care. Future clinical models should incorporate systematic screening for nociplastic pain features, fatigue, and mood disturbance alongside traditional disease activity assessments. Neuroimaging biomarkers, including baseline thalamic and insula-ACC connectivity patterns, may eventually guide treatment stratification by identifying patients most likely to benefit from biologic therapies or adjunctive CNS-targeted interventions.[38,39,53,54] Such an integrated, mechanism-based approach shifts the paradigm from treating to target to treating to recovery, with the ultimate goal of reducing pain, improving function, and enhancing quality of life.

Putting the brain on the rheumatology roadmap

The recognition that CNS mechanisms play a fundamental role in pain, fatigue, and mood disturbance in inflammatory arthritis represents a paradigm shift for rheumatology. Neuroimaging has moved this field forward by providing objective, in vivo markers of central pain processing and demonstrating that brain changes both respond to, and sometimes predict, treatment effects. To realize the clinical potential of these insights, several key directions are now needed.

Larger, longitudinal cohort studies. Most existing neuroimaging studies in inflammatory arthritis have been small, single-center investigations. Larger, longitudinal cohorts will be essential to validate early findings and determine the natural history of CNS changes in relation to disease activity, treatment, and patient-reported outcomes. Such studies could clarify whether specific neuroimaging signatures are stable traits, modifiable states, or both and whether they can predict long-term functional outcomes.[23,36]

Neuroimaging as a stratification tool for clinical trials. There is growing interest in using neuroimaging biomarkers to stratify patients in randomized controlled trials. Baseline thalamic activation and insula-ACC connectivity, for example, have been linked to differential responses to TNF inhibitors.[37,54] Incorporating such markers into trial design could help identify the patients most likely to benefit from specific biologics or adjunctive centrally acting therapies, increasing both the efficiency and success of trials.

Back-translation of mechanistic insights to clinical measures. To ensure that mechanistic insights reach the bedside, there is a need to translate neuroimaging findings into accessible clinical tools. This could include the refinement of composite indices that integrate measures of nociplastic pain features, fatigue, and mood disturbance, informed by brain-based evidence. Routine use of validated patient-reported outcome measures, alongside clinical and laboratory indices, may allow rheumatologists to better identify patients with centrally mediated symptoms, even when advanced imaging is not available.[52,53]

The future of mechanism-based rheumatology. Ultimately, the integration of brain-focused research into rheumatology practice promises a more comprehensive, mechanism-based approach to patient care. By recognizing the CNS as both a target and a mediator of symptom burden, the field can move beyond treating inflammation alone to improving holistic outcomes. Neuroimaging stands to play a pivotal role in this transformation, guiding patient stratification, informing personalized therapy, and closing the gap between inflammatory control and meaningful recovery.

ACKNOWLEDGMENT

Assistance from a large language model (ChatGPT, OpenAI, and Mayo Clinic AI assistant) was used in the preparation of this manuscript for refining phrasing and improving clarity of expression.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Garrido-Cumbrera M, Hillmann O, Mahapatra R, et al. Improving the management of psoriatic arthritis and axial spondyloarthritis:

roundtable discussions with healthcare professionals and patients. Rheumatol Ther 2017;4(2):219–231.
2. ten Klooster PM, Veehof MM, Taal E, et al. Changes in priorities for improvement in patients with rheumatoid arthritis during 1 year of anti-tumour necrosis factor treatment. Ann Rheum Dis 2007;66(11): 1485–1490.
3. Altawil R, Saevarsdottir S, Wedrén S, et al. Remaining pain in early rheumatoid arthritis patients treated with methotrexate. Arthritis Care Res (Hoboken) 2016;68(8):1061–1068.
4. Glinatsi D, Baker JF, Hetland ML, et al. Magnetic resonance imaging assessed inflammation in the wrist is associated with patient-reported physical impairment, global assessment of disease activity and pain in early rheumatoid arthritis: longitudinal results from two randomised controlled trials. Ann Rheum Dis 2017;76(10):1707–1715.
5. Andersson MLE, Forslind K, Hafström I; BARFOT Study Group. Patients with early rheumatoid arthritis in the 2000s have equal disability and pain despite less disease activity compared with the 1990s: data from the BARFOT study over 8 years. J Rheumatol 2017;44(6): 723–731.
6. Buch MH, Eyre S, McGonagle D. Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. Nat Rev Rheumatol 2021;17(1):17–33.
7. de Hair MJH, Jacobs JWG, Schoneveld JLM, et al. Difficult-to-treat rheumatoid arthritis: an area of unmet clinical need. Rheumatology (Oxford) 2018;57(7):1135–1144.
8. Nagy G, Roodenrijs NMT, Welsing PM, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2021;80(1): 31–35.
9. Paudel ML, Li R, Naik C, et al. Prevalence and characteristics of adults with difficult-to-treat rheumatoid arthritis in a large patient registry. Rheumatology (Oxford) 2025;64(3):1102–1110.
10. Sunzini F, Schrepf A, Clauw DJ, et al. The biology of pain: through the rheumatology lens. Arthritis Rheumatol 2023;75(5):650–660.
11. Murphy AE, Minhas D, Clauw DJ, et al. Identifying and managing nociplastic pain in individuals with rheumatic diseases: a narrative review. Arthritis Care Res (Hoboken) 2023;75(10):2215–2222.
12. Nicholas M, Vlaeyen JWS, Rief W, et al; IASP Taskforce for the Classification of Chronic Pain. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain 2019;160(1):28–37.
13. Kosek E, Clauw D, Nijs J, et al. Chronic nociplastic pain affecting the musculoskeletal system: clinical criteria and grading system. Pain 2021;162(11):2629–2634.
14. Kosek E, Cohen M, Baron R, et al. Do we need a third mechanistic descriptor for chronic pain states? Pain 2016;157(7):1382–1386.
15. Rutter-Locher Z, Arumalla N, Norton S, et al. A systematic review and meta-analysis of questionnaires to screen for pain sensitisation and neuropathic like pain in inflammatory arthritis. Semin Arthritis Rheum 2023;61:152207.
16. Kelleher EM, Lange F, Wanigasekera V, et al. Brain signatures of nociplastic pain: Fibromyalgia Index and descending modulation at population level. Brain. Published online August 17, 2025. https://doi.org/10.1093/brain/awaf307
17. Lopes F, Vicentini FA, Cluny NL, et al. Brain TNF drives post-inflammation depression-like behavior and persistent pain in experimental arthritis. Brain Behav Immun 2020;89:224–232.
18. Schaible HG. Nociceptive neurons detect cytokines in arthritis. Arthritis Res Ther 2014;16(5):470.
19. Loggia ML, Segerdahl AR, Howard MA, et al. Imaging clinically relevant pain states using arterial spin labeling. Pain Rep 2019;4(4):e750.
20. Lee MC, Tracey I. Imaging pain: a potent means for investigating pain mechanisms in patients. Br J Anaesth 2013;111(1):64–72.
21. Lee MC, Tracey I. Unravelling the mystery of pain, suffering, and relief with brain imaging. Curr Pain Headache Rep 2010;14(2):124–131.
22. Albrecht DS, Forsberg A, Sandström A, et al. Brain glial activation in fibromyalgia - a multi-site positron emission tomography investigation. Brain Behav Immun 2019;75:72–83.
23. Basu N, Kaplan CM, Ichesco E, et al. Neurobiologic features of fibromyalgia are also present among rheumatoid arthritis patients. Arthritis Rheumatol 2018;70(7):1000–1007.
24. Lee YC, Bingham CO III, Edwards RR, et al. Association between pain sensitization and disease activity in patients with rheumatoid arthritis: a cross-sectional study. Arthritis Care Res (Hoboken) 2018;70(2): 197–204.
25. Heisler AC, Song J, Dunlop DD, et al. Association of pain centralization and patient-reported pain in active rheumatoid arthritis. Arthritis Care Res (Hoboken) 2020;72(8):1122–1129.
26. Sunzini F, Stefanov K, Al-Wasity S, et al. The insula represents a key neurobiological pain hub in psoriatic arthritis. Arthritis Res Ther 2025; 27(1):70.
27. Flodin P, Martinsen S, Altawil R, et al. Intrinsic brain connectivity in chronic pain: a resting-state fMRI study in patients with rheumatoid arthritis. Front Hum Neurosci 2016;10:107.
28. Sandström A, Ellerbrock I, Jensen KB, et al. Altered cerebral pain processing of noxious stimuli from inflamed joints in rheumatoid arthritis: an event-related fMRI study. Brain Behav Immun 2019;81:272–279.
29. Wu Q, Inman RD, Davis KD. Neuropathic pain in ankylosing spondylitis: a psychophysics and brain imaging study. Arthritis Rheum 2013; 65(6):1494–1503.
30. Freynhagen R, Baron R, Gockel U, et al. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006;22(10):1911–1920.
31. Finnerup NB, Haroutounian S, Kamerman P, et al. Neuropathic pain: an updated grading system for research and clinical practice. Pain 2016;157(8):1599–1606.
32. Soni A, Wanigasekera V, Mezue M, et al. Central sensitization in knee osteoarthritis: relating presurgical brainstem neuroimaging and PainDETECT-based patient stratification to arthroplasty outcome. Arthritis Rheumatol 2019;71(4):550–560.
33. Rifbjerg-Madsen S, Wæhrens EE, Danneskiold-Samsøe B, et al. Psychometric properties of the painDETECT questionnaire in rheumatoid arthritis, psoriatic arthritis and spondyloarthritis: Rasch analysis and test-retest reliability. Health Qual Life Outcomes 2017;15(1):110.
34. Hochman JR, Davis AM, Elkayam J, et al. Neuropathic pain symptoms on the modified painDETECT correlate with signs of central sensitization in knee osteoarthritis. Osteoarthritis Cartilage 2013;21(9): 1236–1242.
35. Rehm SE, Koroschetz J, Gockel U, et al. A cross-sectional survey of 3035 patients with fibromyalgia: subgroups of patients with typical comorbidities and sensory symptom profiles. Rheumatology (Oxford) 2010;49(6):1146–1152.
36. Schrepf A, Kaplan CM, Ichesco E, et al. A multi-modal MRI study of the central response to inflammation in rheumatoid arthritis. Nat Commun 2018;9(1):2243.
37. Rech J, Hess A, Finzel S, et al. Association of brain functional magnetic resonance activity with response to tumor necrosis factor inhibition in rheumatoid arthritis. Arthritis Rheum 2013;65(2):325–333.
38. Hess A, Axmann R, Rech J, et al. Blockade of TNF-α rapidly inhibits pain responses in the central nervous system. Proc Natl Acad Sci USA 2011;108(9):3731–3736.
39. Hess A, Tascilar K, Schenker HM, et al. Disease-associated brain activation predicts clinical response to TNF inhibition in rheumatoid arthritis (PreCePra): a randomised, multicentre, double-blind, placebo-controlled phase 3 study. Lancet Rheumatol 2025;7(8):e565–e575.
40. Mueller C, Lin JC, Thannickal HH, et al. No evidence of abnormal metabolic or inflammatory activity in the brains of patients with rheumatoid arthritis: results from a preliminary study using whole-brain magnetic

resonance spectroscopic imaging (MRSI). Clin Rheumatol 2020; 39(6):1765–1774.
41. Lee YC, Fine A, Protsenko E, et al. Brain correlates of continuous pain in rheumatoid arthritis as measured by pulsed arterial spin labeling. Arthritis Care Res (Hoboken) 2019;71(2):308–318.
42. Kaplan C, Minc A, Basu N, et al. Inflammation and the central nervous system in inflammatory rheumatic disease. Curr Rheumatol Rep 2019;21(12):67.
43. Capuron L, Miller AH. Cytokines and psychopathology: lessons from interferon-alpha. Biol Psychiatry 2004;56(11):819–824.
44. Capuron L, Raison CL, Musselman DL, et al. Association of exaggerated HPA axis response to the initial injection of interferon-alpha with development of depression during interferon-alpha therapy. Am J Psychiatry 2003;160(7):1342–1345.
45. Raison CL, Borisov AS, Broadwell SD, et al. Depression during pegylated interferon-alpha plus ribavirin therapy: prevalence and prediction. J Clin Psychiatry 2005;66(1):41–48.
46. Raison CL, Broadwell SD, Borisov AS, et al. Depressive symptoms and viral clearance in patients receiving interferon-alpha and ribavirin for hepatitis C. Brain Behav Immun 2005;19(1):23–27.
47. Raison CL, Demetrashvili M, Capuron L, et al. Neuropsychiatric adverse effects of interferon-alpha: recognition and management. CNS Drugs 2005;19(2):105–123.
48. Capuron L, Gumnick JF, Musselman DL, et al. Neurobehavioral effects of interferon-alpha in cancer patients: phenomenology and paroxetine responsiveness of symptom dimensions. Neuropsychopharmacology 2002;26(5):643–652.
49. Capuron L, Pagnoni G, Demetrashvili M, et al. Anterior cingulate activation and error processing during interferon-alpha treatment. Biol Psychiatry 2005;58(3):190–196.
50. Capuron L, Pagnoni G, Demetrashvili MF, et al. Basal ganglia hypermetabolism and symptoms of fatigue during interferon-alpha therapy. Neuropsychopharmacology 2007;32(11):2384–2392.
51. Capuron L, Pagnoni G, Drake DF, et al. Dopaminergic mechanisms of reduced basal ganglia responses to hedonic reward during interferon alfa administration. Arch Gen Psychiatry 2012;69(10):1044–1053.
52. Kaplan CM, Schrepf A, Ichesco E, et al. Association of inflammation with pronociceptive brain connections in rheumatoid arthritis patients with concomitant fibromyalgia. Arthritis Rheumatol 2020;72(1):41–46.
53. Wartolowska K, Jenkinson M, Andersson J, et al. Imaging the effects of tumour necrosis factor inhibition on pain processing in severe rheumatoid arthritis. medRxiv Preprint posted online March 1, 2023. https://doi.org/10.1101/2023.02.27.23286245
54. Abe N, Fujieda Y, Tha KK, et al. Aberrant functional connectivity between anterior cingulate cortex and left insula in association with therapeutic response to biologics in inflammatory arthritis. Semin Arthritis Rheum 2022;55:151994.
55. Wu Q, Inman RD, Davis KD. Tumor necrosis factor inhibitor therapy in ankylosing spondylitis: differential effects on pain and fatigue and brain correlates. Pain 2015;156(2):297–304.
56. Moayedi M, Weissman-Fogel I, Crawley AP, et al. Contribution of chronic pain and neuroticism to abnormal forebrain gray matter in patients with temporomandibular disorder. Neuroimage 2011;55(1): 277–286.
57. McCrae CS, Mundt JM, Curtis AF, et al. Gray matter changes following cognitive behavioral therapy for patients with comorbid fibromyalgia and insomnia: a pilot study. J Clin Sleep Med 2018; 14(9):1595–1603.
58. Zhang A, Lee YC. Mechanisms for joint pain in rheumatoid arthritis (RA): from cytokines to central sensitization. Curr Osteoporos Rep 2018;16(5):603–610.
59. Taylor PC. Pain in the joints and beyond; the challenge of rheumatoid arthritis. Lancet Rheumatol 2023;5(6):e351–e360.
60. Taylor PC, Keystone EC, van der Heijde D, et al. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. N Engl J Med 2017; 376(7):652–662.
61. Dougados M, Taylor PC, Bingham CO III, et al. The effect of tofacitinib on residual pain in patients with rheumatoid arthritis and psoriatic arthritis. RMD Open 2022;8(2):e002478.
62. Espartal E, Michelena X, Marsal S, et al. POS1605 A proof-of-concept study evaluating the use of functional brain magnetic resonance imaging in assessing treatment response in psoriatic arthritis patients. Ann Rheum Dis 2023;82:1147.
63. Baliki M, Katz J, Chialvo DR, et al. Single subject pharmacological-MRI (phMRI) study: modulation of brain activity of psoriatic arthritis pain by cyclooxygenase-2 inhibitor. Mol Pain 2005;1:32.
64. Wartolowska K, Tracey I. Neuroimaging as a tool for pain diagnosis and analgesic development. Neurotherapeutics 2009;6(4):755–760.
65. Schweinhardt P, Kalk N, Wartolowska K, et al. Investigation into the neural correlates of emotional augmentation of clinical pain. Neuroimage 2008;40(2):759–766.
66. Hua K, Wang P, Xia B, et al. Pain- and depression-related regional homogeneity changes in ankylosing spondylitis: a functional magnetic resonance imaging study. Front Neurol 2025;16:1521531.
67. Liu Q, Liao Z, Zhang Y, et al. Pain- and fatigue-related functional and structural changes in ankylosing spondylitis: an fRMI study. Front Med (Lausanne) 2020;7:193.
68. Hua K, Wang P, Lan Z, et al. Increased left putamen volume correlates with pain in ankylosing spondylitis patients. Front Neurol 2020;11: 607646.
69. Passia E, Vis M, Coates LC, et al. Sex-specific differences and how to handle them in early psoriatic arthritis. Arthritis Res Ther 2022; 24(1):22.
70. Kim JR, Kim HA. Molecular mechanisms of sex-related differences in arthritis and associated pain. Int J Mol Sci 2020;21(21):7938.
71. Vogel K, Muhammad LN, Song J, et al. Sex differences in pain and quantitative sensory testing in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2023;75(12):2472–2480.
72. Fauchon C, Meunier D, Rogachov A, et al. Sex differences in brain modular organization in chronic pain. Pain 2021;162(4):1188–1200.
73. Rogachov A, Cheng JC, Hemington KS, et al. Abnormal low-frequency oscillations reflect trait-like pain ratings in chronic pain patients revealed through a machine learning approach. J Neurosci 2018;38(33):7293–7302.
74. Rogachov A, Cheng JC, Erpelding N, et al. Regional brain signal variability: a novel indicator of pain sensitivity and coping. Pain 2016; 157(11):2483–2492.
75. Hemington KS, Rogachov A, Cheng JC, et al. Patients with chronic pain exhibit a complex relationship triad between pain, resilience, and within- and cross-network functional connectivity of the default mode network. Pain 2018;159(8):1621–1630.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 15–30
DOI 10.1002/acr.25532

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

Longitudinal Relationship Between Pain and Depression in People With Inflammatory Arthritis: A Narrative Review

Natasha Cox,[1] Saeed Farooq,[2] Helen Twohig,[2] and Ian C. Scott[1]

As many patients with inflammatory arthritis (IA) have chronic pain, understanding how to best assess and manage pain in IA is a priority. Comorbid depression is prevalent in adults with IA, affecting 15% to 39% of people. Although pain and depression are thought to be associated in IA, this concept is largely based on cross-sectional data. To better understand potential causality, longitudinal studies are required. This narrative review examines the relationship between pain and depression in patients with IA participating in observational longitudinal studies (spanning association strengths, direction of effect, and causal factors) and summarizes the literature on causal pathways in general populations alongside guideline recommendations/systematic reviews on assessing pain/depression in IA. Fourteen longitudinal studies in IA largely indicate an association between pain and depression, albeit with a small-to-modest effect size and a focus on pain intensity. This relationship appears to be bidirectional. Evidence on causal pathways is sparse in IA and limited in non-IA populations (with structural/function brain changes, altered sensory/pain thresholds, and sleep disturbance implicated) highlighting a need for further research. Although many patient-reported outcome measures exist to assess pain and depression in IA, evidence for their psychometric properties is often limited, and IA guidelines offer incomplete advice on pain/depression assessment. A simple approach of using a single-item pain intensity score (eg, a Numeric Rating Scale, which has strong clinimetric properties) in routine IA consultations, with screening for depression where relevant, appears appropriate. Further research is needed to understand how this could be achieved in different health care settings.

Introduction

Despite substantial therapeutic advances in treating synovitis, chronic pain remains a problem for many patients with inflammatory arthritis (IA).[1] Surveys demonstrate that most patients with IA have daily pain of moderate-to-severe intensity,[2,3] and longitudinal studies demonstrate that many patients with rheumatoid arthritis (RA),[4,5] axial spondylarthritis,[6] and psoriatic arthritis (PsA) receiving biologic disease-modifying antirheumatic drugs experience persistent pain.[7] Chronic pain in IA has widespread impacts, associating with worse quality of life, disability, and fatigue[7–10] alongside high levels of long-term opioid and gabapentinoid prescribing[11] (despite absent trial evidence of efficacy).[12] Consequently, understanding how to best assess and manage pain in IA is a priority.[13]

Depression affects an estimated 15% to 39% of adults with IA, depending on the definition used[14,15] (far higher than the estimated 5% of adults affected in the general population[16]). Comorbid depression in IA has been reported to associate with lesser treatment responses,[17] worse quality of life,[18] and higher mortality.[19] Although widely believed that depression associates with pain in IA, this is largely based on cross-sectional data showing moderate correlations between depression levels and pain intensity scores.[20–23] As longitudinal studies are required to better understand potential causality, the primary aim of this narrative review is to summarize longitudinal observational studies examining the

The views expressed are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Dr Cox's work was supported by the Haywood Foundation and the Keele Haywood Academic Rheumatology Group. Dr Scott's work was supported by the National Institute for Health and Care Research (grant NIHR300826).

[1]Natasha Cox, MBChB, Ian Scott, PhD, MSc, MRCP, MBChB: Primary Care Centre Versus Arthritis at Keele University, Keele, Haywood Academic Rheumatology Centre, Haywood Hospital, Midlands Partnership University National Health Service Foundation Trust, Staffordshire, United Kingdom; [2]Saeed Farooq, PhD, MCPS (Psych), FCPS (Psych), Helen Twohig, PhD, MBChB, BMedSci, MRCP, MRCGP: Primary Care Centre Versus Arthritis at Keele University, Keele, United Kingdom.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25532).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25532.

Address correspondence via email to Natasha Cox, MBChB, at email. n.cox1@keele.ac.uk

Submitted for publication February 28, 2025; accepted in revised form March 11, 2025.

relationship between pain and depression in IA (spanning the strength of associations, direction of effect, and potential causal factors). Its secondary aims are to summarize literature on (a) potential causal pathways underlying the associations between pain and depression in general populations and (b) the best ways to measure and assess pain and depression in routine IA care. This information will be of substantial value to rheumatologists (and other health care professionals, including those in primary care) in informing their use of biopsychosocial IA pain care.

Methods

To address our primary aim, we searched three bibliographic databases (MEDLINE, PsycINFO, and CINAHL) for longitudinal observational studies examining the relationship between depression and pain in the most common types of IA (RA and/or spondyloarthritis), including studies considering the association between pain and depression and/or variables explaining this relationship. For our secondary aims, we searched PubMed for systematic literature reviews examining (a) causal pathways for the relationship between pain and depression in adults published in the last decade and (b) tools to assess pain/depression in IA. Searches were conducted by one author (NC), with an overview of the approach detailed in Supplementary Figure 1.

Association between pain and depression in IA

Overview. This section considers longitudinal study evidence for the association. From 6,505 references, 14 relevant studies were identified (Table 1). All considered pain intensity using a range of analytical methods; two also considered other pain dimensions.

Correlations between depression and pain. These were reported in two studies. We classified correlation coefficients as very strong (0.8–1), strong (0.6–0.8), moderate (0.4–0.6), weak (0.2–0.4), and very weak (0–0.2).[24,25] Gialouri et al examined the course of depression and its possible associations with disease-related variables/patient-reported outcome measures (PROMs) in 93 patients with PsA enrolled from two hospitals. Data were considered at two sequential visits (between 2019 and 2021). The correlation between changes in depression scores (measured using the Hospital Anxiety Depression Scale [HADS]) and changes in pain intensity scores (measured using a Numeric Rating Scale [NRS]) indicated a weak, statistically significant correlation ($r = 0.27$, $P = 0.01$).[26] Chaney et al examined the correlation between baseline pain intensity scores and depression scores at one follow-up timepoint (mean 12.2 months) in 42 patients with RA. They used the 18-item Inventory to Diagnose Depression and Modified Stanford Health Assessment Questionnaire Pain scale, reporting a moderate, significant correlation ($r = 0.53$, $P = 0.01$).[27]

Depression explaining pain. Four studies evaluated the association between depression as an explanatory variable and pain intensity an outcome variable. Three reported statistically significant associations, and one had mixed findings. First, Graham-Engeland et al examined whether baseline depressive symptoms predicted mean momentary pain scores in 31 patients with RA over 7 days using multilevel modeling. Depression was measured at baseline using the Centre for Epidemiological Studies Depression Scale (CES-D) (0–60; ≥19 indicative of depression), and pain was measured five times a day for 7 days (using self-reported scales, ranging from 0 to 6 for stiffness, pain, and joint tenderness/swelling; the mean of these created a "total pain scale"). Higher baseline depression scores associated with higher pain intensity scores (unstandardized $\beta = 0.09$; $P < 0.05$),[28] indicating that for every 10-unit increase in baseline depression scores, a corresponding 0.9-point increase in mean pain scores during follow-up was observed.

Second, Ward assessed the association between depression and pain in 24 patients with RA over 60 weeks (data collected twice-weekly) using a pooled time-series regression. Depression was measured using the CES-D scale (modified version, excluding items closely related to arthritis to ensure depression was not overestimated) and pain using the Health Assessment Questionnaire (HAQ) pain scale (0–3). Change in depression scores significantly associated with change in pain intensity scores (partial $R^2 = 0.06$; $P < 0.0001$), indicating that changes in depression explained 6% of the variation in changes in pain.[29]

Third, Snoeck Henkemans et al examined the association between depressive symptoms and disease activity (including pain) in 367 patients with PsA over 2 years using linear mixed-effects models. Depression was measured using the depression subscale of HADS (0–21; scores >7 indicating depression) and pain using a 10-cm Visual Analog Scale (VAS). Depression at any time during follow-up associated with higher mean pain scores (evidenced by graphical representation, although raw data were not reported).[30] Fourth, Rathbun et al assessed whether depression affects disease activity (including pain) in 268 veterans with early RA prescribed methotrexate. They used weighted generalized estimating equations to model outcomes at 6, 12, and 24 months. Depression was determined by the presence of an *International Classification of Disease, Ninth Revision* (*ICD-9*) code, and pain using a 10-cm VAS. Baseline depression significantly associated with higher pain scores at 6 ($P = 0.029$) and 12 ($P = 0.028$) but not 24 months ($P = 0.735$), suggesting that differences in pain levels by depression status may decline over time. As depression was only measured at baseline, and depression status could have altered over follow-up, this could have biased 2-year findings to the null.[31]

Pain explaining depression. Three studies evaluated the association between pain intensity as an explanatory variable and

Table 1. Longitudinal studies examining the relationship between depression and pain in patients with IA*

Study, year	Region	IA type	Study duration	Size, n	Female, %	Age, years	How the relationship was evaluated	Depression measure	Pain measure	Key findings
Snoeck Henkemans et al,[30] 2024	The Netherlands	PsA	2 years	367	49	49–51	Association between mean pain and depression scores at any point in study	HADS-D	VAS	Higher depression scores associated with higher pain scores
Gialouri et al,[26] 2023	Greece	PsA	29 months	128	52	53	Correlation between changes in depression and pain scores between two clinic visits	HADS-D	PtP	Weak, statistically significant correlation (r = 0.27, *P* = 0.01)
Rathbun et al,[31] 2021	USA	RA	2 years	268	11	59–64	Difference in follow-up pain scores between patients with and without depression at baseline	*ICD-9*	VAS	Depression associated with higher pain scores at 6 months and 1 year but not at 2 years
Graham-Engeland et al,[28] 2016[a]	USA	RA	7 days	31	74	50	Association between baseline depression scores and mean momentary pain scores and the mediating effect of momentary mood on this association	CES-D	Bespoke	Worse depression scores associated with more intense pain (β = 0.09, *P* < 0.05). Momentary mood did not mediate the pain depression association
Husted et al,[35] 2012	Canada	PsA	7.5 years	394	29	45	Association between (a) changes in pain and depression scores between visits (and vice versa) and (b) pain scores at previous visit and changes in depression scores between visits (and vice versa)	SF-36	HAQ	Small bidirectional relationship (standardized regression coefficients < 0.3) between depressive symptoms and pain
McQuillan et al,[33] 2012	USA	RA	10 years	854	83	58	Association between changes in pain and depression scores	CES-D	NRS	Changes in pain scores associated with changes in depression scores (β = 0.04; *P* < 0.001)
Odegård et al,[38]	Norway	RA	10 years	149	74	52	Association between pain and depression scores over time (and vice versa)	AIMS	VAS	No association between depression and pain scores (and vice versa)
Conner et al,[34] 2006[a]	USA	RA	30 days	188	68	56	Association between (a) depression history and mean pain and (b) mean depression and mean pain	DSM-IV	NRS	Depression history was not associated with pain; current depression associated with pain (β = 7.81, *P* < 0.001)
Chaney et al,[27] 2004	USA	RA	1 year	42	81	53	Correlation between baseline pain and follow-up depression scores	IDD	MHAQ-P	Moderate, statistically significant correlation (r = 0.53; *P* = 0.01)

(Continued)

Table 1. (*Cont'd*)

Study, year	Region	IA type	Study duration	Size, n	Female, %	Age, years	How the relationship was evaluated	Depression measure	Pain measure	Key findings
Smedstad et al,[37] 1997	Norway	RA	2 years	216	73	52	Direction of effect of association between changes in depression and pain scores over 1 year	AIMS	VAS	Inconsistent results
Ward,[29] 1994	USA	RA	14 months	24	92	46	Association between changes in depression and pain scores	CES-D	HAQ	Depression associated with pain (partial R^2 = 0.06; $P < 0.0001$)
Wolfe and Hawley,[32] 1993	USA	RA	10 years	713	71	60	Association between changes in pain and depression scores between two clinic visits	AIMS	VAS	Changes in pain scores associated with changes in depression scores
Brown et al,[50] 1989	USA	RA	6 months	287	75	51	Interaction effects between pain and pain-coping strategies on depression	CES-D	AIMS	More use of passive coping and more pain led to more severe depression over time
Brown,[36] 1990	USA	RA	3.5 years	243	75	53	Association (including direction) over 6 months assessed by causal modeling	CES-D	VAS	Pain predicted depression in last 6 months; depression did not predict pain
Hawley and Wolfe,[39] 1988	USA	RA	3.1 years	400[b]	74	55	Association among (a) mean depression and mean pain scores, (b) baseline depression and follow-up pain scores, and (c) pain and improvement/worsening in depression scores	AIMS	VAS	(a) Mean depression scores associated with mean pain scores, (b) higher baseline depression scores associated with higher follow-up pain scores, and (c) pain scores did not associate with improvement/worsening in depression scores

* Age is noted as mean, median, or range. AIMS, Arthritis Impact Measurement Scale; CES-D, Center for Epidemiological Studies Depression; DSM-IV, *Diagnostic and Statistical Manual, Fourth Edition*; HADS-D, Hospital Anxiety Depression Scale-Depression subscale; HAQ, Health Assessment Questionnaire; IA, inflammatory arthritis; *ICD-9, International Classification of Diseases, Ninth Edition*; IDD, Inventory to Diagnose Depression; MHAQ-P, Modified Health Assessment Questionnaire Pain; NRS, Numeric Rating Scale; PsA, psoriatic arthritis; PtP, Patient Pain Assessment; RA, rheumatoid arthritis; SF-36, Short Form 36; VAS, Visual Analog Scale.

[a] Also considered pain appraisal and coping (Conner et al[34]) and pain-related restrictions (Graham-Engeland et al[28]).

[b] Study included mixed population but only findings from patients with IA were considered.

depression an outcome variable. Two reported statistically significant associations and one had mixed findings. First, Wolfe and Hawley examined the extent to which clinical variables (including pain) explained depression scores between two clinic visits in 713 patients with RA using stepwise regression models. The Arthritis Impact Measurement Scale (AIMS) depression subscale (0–10) measured depression, and a VAS (0–3) measured pain. Change in pain scores significantly associated with change in depression scores (partial $R^2 = 0.04$; $P < 0.001$), suggesting that 4% of the change in depression scores could be explained by the change in pain scores.[32] As most patients had visits <1 year apart, the longer-term applicability of these findings is uncertain.

Second, McQuillan et al examined associations between changes in RA symptoms (including pain) and depressive symptoms in 854 patients with RA over 10 years (data collected annually) using multilevel modeling. The CES-D measured depression and an NRS (0–100) measured pain. Changes in pain scores associated with changes in depressive symptoms (unstandardized β = 0.04; $P < 0.001$),[33] indicating that for every 10-unit increase in pain scores a 0.4 increase in depression scores was seen.

Third, Conner et al examined the association between depression (as both a binary and continuous variable) and pain intensity in 188 patients over 30 days (data collected using daily diaries). A history of depression was assessed by interview and pain intensity using an NRS (0–100). Using multilevel modeling they reported that, although depression history did not associate with mean daily pain scores, current depression did (unstandardized β = 7.81; $P < 0.001$) indicating that for every additional depressive symptom, a 7.81-unit increase in pain was observed.[34]

Bidirectional relationship between pain and depression. Five studies evaluated this using different approaches. One reported statistically significant associations in both directions, two reported mixed findings, and two did not report significant associations. First, Husted et al examined whether depressive symptoms influenced changes in pain scores and vice versa in 394 patients with PsA over 7.5 years using linear mixed-effects models. Depression was measured using the Short Form 36 (SF-36) mental component score and pain using the HAQ pain scale. Although the strongest predictors of changes in pain and depressive symptoms between visits were their scores at the previous visit (standardized regression coefficients >0.75) there was evidence of a small bidirectional relationship (standardized regression coefficients <0.3) between depressive symptoms and pain.[35]

Second, Brown examined the extent to which chronic pain and depression were associated in 243 patients with RA, collecting data every 6 months for seven waves (first and final 12 months used for analysis). Depression was measured using the CES-D and pain using the AIMS pain subscale and 10-cm pain VAS. Using structural equation modeling, pain scores predicted depression scores over 6 months in the final 12 months (but not the initial 12 months); depression scores did not predict pain scores.[36]

Third, Smedstad et al explored the bidirectional relationship between depression and pain in 216 patients with RA over 24 months. Using multiple linear regression models and a cross-lagged panel approach they assessed whether changes in depression (AIMS depression scale) influenced subsequent pain levels (100-mm VAS) and vice versa. Depression scores at one timepoint did not predict pain 1 year later (β = 0.00; P = "not significant" at 12 months; β = 0.03; P = "not significant" at 24 months). Similarly, pain scores did not predict future depression (β = –0.12; P = "not significant" at 12 months; β = 0.06; P = "not significant" at 24 months).[37]

Fourth, Odegård et al examined the association between pain and depression in 149 patients with RA over 10 years (follow-up data collected at 1, 2, 5, and 10 years) using a repeated-measures mixed model. The AIMS depression scale and 100-mm pain VAS were used. They observed no significant association between depression and the course of pain and vice versa (statistical values not provided).[38]

Fifth, Hawley and Wolfe, in 400 patients with RA over 4.5 years (data capture every 6 months) used linear regression models to explore the association between mean depression scores and mean pain scores and baseline depression scores and follow-up pain scores. They also used logistic regression models to explore the association between baseline pain scores and improvements/worsening in depression scores. The AIMS depression and HAQ pain subscales were used. Mean depression scores associated with mean pain scores (β = 0.21; $P < 0.001$)[39] and higher baseline depression scores associated with higher follow-up pain scores (β = not reported; $P = 0.047$). Pain did not associate with clinically significant improvements/worsening in depression scores (statistical values not provided).[39]

Depression and other pain dimensions. Two studies also considered other pain dimensions. Graham-Engeland et al used multivariable regression modeling to explore the impact of depression at study baseline on mean "pain-related restrictions" over 1 week (0–6 scale). Patients with depression had higher levels of pain-related restrictions (unstandardized β = 0.08; $P < 0.01$).[28] Conner et al used multilevel modeling to assess whether pain "coping strategies" and "coping appraisals" associated with depression.[34] Pain-coping strategies were assessed using seven items, with participants asked if they used/thought of various strategies to help with their arthritis pain (eg, relaxation/distraction). Pain-coping appraisals were assessed using Likert scales and included catastrophizing and positive appraisals. They reported mixed results, with four (of seven) coping strategies (comprising reappraisal, venting emotions, spiritual comfort, and emotional support) and one (of three) coping appraisals (catastrophizing) associating with depression.[34]

Summary. These longitudinal studies (with a few exceptions) support cross-sectional studies demonstrating an association between pain and depression in patients with IA. They also suggest a bidirectional relationship; although among the five studies considering the association from both directions there was only a statistically significant bidirectional effect in one study, among the four studies examining depression as an explanatory variable for pain and the three studies examining pain as an explanatory variable for depression, all demonstrated positive and statistically significant associations in some form. This has several caveats. First, the strength of the associations (and corresponding effect sizes) is small to modest. Second, most evidence is for a single dimension (intensity) of a multidimensional symptom (pain). Third, the evidence for this relationship is strongest in short- to medium-duration studies but more conflicting in longer-term studies.

Potential causal pathways

Overview. This section begins by considering pain and depression definitions/mechanisms. It then describes evidence from longitudinal studies examining mediators of the association between pain and depression in IA, followed by systematic literature reviews examining causal pathways in non-IA populations.

Pain definition and mechanisms. The International Association for the Study of Pain defines pain as "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage."[40,41] This definition was recently updated to include "key notes." Of particular relevance to depression is the note that pain "is always a personal experience that is influenced to varying degrees by biologic, psychological, and social factors."[41] Pain is conceptually classified into three main categories: nociceptive pain (from actual or threatened damage to nonneural tissue, with nociceptor activation), neuropathic pain (from a lesion/disease of the somatosensory nervous system), and nociplastic pain (from altered nociception despite no clear evidence of actual/threatened tissue damage causing peripheral nociceptor activation or evidence for a disease/lesion of the somatosensory system).[42]

Traditionally, pain in IA was thought to be dominated by nociceptor activation from synovial inflammation and joint damage. However, the roles of nociplastic and neuropathic mechanisms are increasingly recognized,[43,44] with a recent systematic review reporting the prevalence of pain sensitivity/neuropathic-like pain in patients with IA to be 36% to 42%[45] and comorbid fibromyalgia estimated to affect 21%, 18%, and 13% of patients with RA, PsA, and ankylosing spondylitis, respectively.[46] Consequently, it has been proposed that pain in IA is a "mixed type," defined as "a complex overlap of the different known pain types (nociceptive, neuropathic, [and] nociplastic) in any combination, acting simultaneously and/or concurrently to cause pain in the same body area."[47] This complexity complicates IA pain research and treatment.

Depression definition and mechanisms. Two criteria are often used to define depression: *ICD-11* and *Diagnostic Statistical Manual of Mental Health Disorders, Fifth Edition*, which use the terms "depressive episode" and "major depressive episode," respectively. Both require at least five (from nine) characteristic depressive symptoms to be present during the same 2-week period accompanied by a change from normal functioning.

Depression's underlying mechanisms remain unclear, with various factors considered important, including monoamine deficiency, genetic/environmental factors, neurogenesis, and immunologic/endocrinological abnormalities. Hypothalamic pituitary adrenal axis dysregulation is considered to provide a neurobiological link between many of these factors and depression.[48] A recent narrative review has examined factors that may underlie the increased prevalence of comorbid depression in IA, with shared cytokine pathways, inflammatory mediated neurologic changes, and the psychological and social impacts of chronic disease considered potentially important.[49]

Longitudinal studies of causal factors in IA. From 6,505 references, we identified two studies exploring mediators of the association between pain and depression in patients with IA (one was also included in the analysis of associations). First, Brown et al examined the role of coping strategies during pain episodes in predicting depression in 287 patients with RA using cross-lagged, covariance structural modeling. They conducted two data collection "waves" 2 months apart. They reported that "passive" coping strategies (eg, depending on others/restricting activities) associated with higher levels of depression. Conversely, "active" coping strategies (eg, staying busy/ignoring pain) had lesser associations with depression. They also reported that an interaction between high pain levels and frequent use of passive coping strategies resulted in the greatest levels of depression. Conversely, active coping strategies had an inverse association with depression.[50]

Second, Graham-Engeland et al, previously described, examined whether baseline depressive symptoms and momentary mood fluctuations (both positive and negative) predicted momentary pain among 31 adults with RA using multilevel modeling. They measured depression at baseline using the CES-D, with momentary pain and mood data collected five times a day for 1 week. Baseline depressive symptoms associated with higher levels of momentary pain and pain-related restrictions. Although greater momentary positive mood associated with less pain and fewer pain-related restrictions, greater momentary negative mood associated with more pain-related restrictions but did not significantly predict pain after controlling for stress experiences.[28]

Systematic reviews of causal pathways in non-IA populations. From 729 potential references, nine relevant systematic reviews were identified (Table 2). Causal pathways considered can be classified as neurobiologic (five reviews), psychosocial (one review), and multiple mechanisms (three reviews).

Neurobiologic. The emergence of neuroimaging, particularly magnetic resonance imaging (MRI), has provided insights into brain structural/functional changes (termed "neuroplasticity") occurring in chronic pain.[60] Four reviews synthesized literature on brain regions/morphologic brain changes occurring in association with pain and depression. First, Malfliet et al evaluated the relationship between maladaptive cognitive and emotional factors related to pain (including depression) and brain alterations (measured by MRI/functional MRI [fMRI]) in patients with chronic pain. A total of 14 included studies considered depression (measured by self-reported questionnaires), observing evidence for the following MRI changes: (1) an inverse relationship with resting state functional connectivity in brain areas involved in pain processing (one case-control study in patients with fibromyalgia, considered "weak" methodologic quality), (2) a positive correlation with parts of the default mode network involved in affective-cognitive processing (one case-control study in patients with localized provoked vulvodynia, considered "strong" methodologic quality), and (3) a positive association with pain-induced activation of the left prefrontal cortex (one case-control study in patients with irritable bowel syndrome, considered "weak" methodologic quality). However, many studies reported no associations between depression and brain changes, variation in MRI techniques across studies limited the interpretation of the findings, and it is unknown if these changes also occur in people without chronic pain.[54]

Second, Zheng et al used activation likelihood estimation meta-analysis to identify neural correlations with comorbid pain and depression. Activation likelihood estimation meta-analysis represents a quantitative approach for pooling neuroimaging metadata.[61] Twenty-six studies using MRI, fMRI, or positron emission tomography were included (most considered "poor"/"fair" quality). Meta-analysis suggested that, in people with pain as the primary diagnosis, concomitant depression associated with increased activation of the right parahippocampal gyrus/amygdala, whereas in people with depression as the primary diagnosis, concomitant pain associated with increased activation of the left superior frontal gyrus and left thalamus. However, only studies reporting statistically significant results were included, with many studies reporting no significant associations.[52]

Third, as amygdala size has been reported to increase in people with major depression, Chen et al synthesized literature on the relationship between emotional stress and MRI-measured amygdala size in patients with chronic pain versus healthy controls. Emotional stress was captured using outcomes considering both depression and anxiety. Thirteen studies were included; seven reported significant amygdala shrinkage in patients with chronic pain and emotional stress versus controls (Hedges g = −0.73; $P = 0.02$). However, it cannot be established whether these changes are related to depression, anxiety, chronic pain or confounding variables.[51]

Fourth, as gray matter volume alterations are considered key neuroplastic changes in people with chronic pain and depressive symptoms, Ma et al synthesized data from relevant voxel-based morphometry studies (a method to analyze brain microstructural changes) comparing gray matter changes in patients with chronic pain and associated depression with healthy controls. Eighteen studies were included. Compared with healthy controls, people with chronic pain and depression had increased gray matter volume in the left hippocampus ($P < 0.005$) and decreased gray matter volume in the medial part of the left superior frontal gyrus ($P < 0.005$).[53]

One systematic review explored responses to painful stimuli according to the presence of depression, with Thompson et al conducting a meta-analysis to compare responses to experimentally induced pain in depressed and healthy groups.[55] A total of 32 studies were included. Outcomes considered were pain threshold (when pain is first perceived), tolerance (when pain can no longer be tolerated), and intensity. Pain tolerance for high-intensity noxious stimulation was similar between groups (Hedges g = 0.09, $P = 0.71$; 10 studies), but for low-intensity stimulation a small, statistically significant higher mean sensory threshold (g = 0.35, $P = 0.01$; nine studies) and pain threshold (g = 0.32, $P = 0.02$; 25 studies) was observed in depressed versus control groups. Although this finding suggests that people with depression may have a higher pain threshold for low-intensity noxious stimuli, the effect size was small and marked heterogeneity was observed. A proposed conceptual model for this paradoxical finding was that, for low-intensity painful stimuli, greater attention is devoted to nonpainful stimuli (eg, internal thoughts) reducing mild pain by denying it attention resources but having no impact for higher-intensity painful stimuli demanding greater attention.

Psychosocial. One review considered psychosocial pathways, with Karimi et al examining the extent to which sleep disturbance mediates the relationship between depression and chronic pain. A total of 49 studies were included, with most (37 studies) considered to be of high quality. They reported, with moderate certainty in the evidence, that 12.5% of the total effect of depression on pain could be accounted for by the indirect effect of disturbed sleep, although causality could not be inferred.[56]

Multiple mechanisms. Three reviews considered both biologic and psychosocial pathways. First, Antoniou et al synthesized neuroimaging study findings to examine the impact of adverse childhood experiences on neural correlations with chronic pain and depression (either alone or in combination). From 49 included studies, 43 considered adverse childhood experience and major depressive disorder and 4 considered adverse childhood experience and chronic pain (no studies

Table 2. Systematic reviews of potential causal pathways for the association between pain and depression in noninflammatory arthritis populations*

Study, year	Aim	Key findings	Limitation(s)
Neurobiologic mechanisms			
Chen et al,[51] 2023	Examine relationship between emotional stress and amygdala size in patients with chronic pain and emotional stress	Thirteen studies synthesized. Seven indicated significant amygdala shrinkage (measured using MRI) in patients with chronic pain and emotional stress compared with healthy controls ($P = 0.02$)	Not possible to compare amygdala size in those with chronic pain with/without emotional distress because of insufficient studies
Zheng et al,[52] 2022	Identify neural correlates of comorbid pain and depression, using ALE meta-analysis	Twenty-six studies combined in the meta-analysis. Results indicated that pain with concomitant depression associated with the right amygdala, whereas depression with concomitant pain related to the left dorsolateral prefrontal cortex	Only included studies in meta-analysis with ALE coordinates corresponding to significant results
Ma et al,[53] 2022	Collect VBM studies on gray mater volume alterations comparing patients with chronic pain and associated depression to healthy controls to explore brain regions involved in comorbid chronic pain and depression	Eighteen studies included. Compared with healthy controls, people with chronic pain and depression had increased gray matter volume in the left hippocampus ($P < 0.005$) and decreased gray matter volume in the medial part of the left superior frontal gyrus ($P < 0.005$)	The comparison group was healthy controls and people with chronic pain but not depression, or people with depression without chronic pain
Malfliet et al,[54] 2016	Synthesize literature on the relationship between maladaptive cognitive and emotional factors related to pain and brain alterations in chronic pain using MRI	Fourteen articles considered depressive symptoms. Evidence for the following in relation to depression: (a) inverse relationships with resting state functional connectivity in brain areas involved in pain processing, (b) positive correlation with parts of the default mode network involved in affective-cognitive processing, and (c) positive association with pain-induced activation of the left prefrontal cortex	MRI processing varied, making interpretation difficult; unknown if changes also occur in people without chronic pain
Thompson et al,[55] 2016	Conduct a meta-analysis to compare depressed and healthy control groups and their response to experimentally induced pain	Thirty-two studies included in meta-analysis. Comparing depressed and healthy groups, for high-intensity noxious stimulation, overall pain tolerance was similar. For low-intensity stimulation, depressed groups had slightly higher mean sensory threshold and pain threshold	Significant heterogeneity present
Psychosocial mechanisms			
Karimi et al,[56] 2023	Examine the extent to which sleep disturbance mediates the association between depression and chronic pain	Forty-nine studies included. A total of 12.5% of the total effect of depression on pain could be explained by the indirect effect of sleep disturbance (moderate certainty in evidence)	Significant heterogeneity was present
Multiple mechanisms			
Antoniou et al,[57] 2023	Investigate impact of adverse childhood experiences on the neural correlates of chronic pain and depression, using neuroimaging studies	Forty-nine studies included. Compared with those without adverse childhood experiences, those with depression and adverse childhood experiences had functional and structural brain changes in the superior frontal, lingual gyrus, hippocampus, insula, putamen, superior temporal, inferior temporal gyrus, and anterior cerebellum and those with depression or chronic pain	No studies considered people with both chronic pain and depression

(Continued)

Table 2. (*Cont'd*)

Study, year	Aim	Key findings	Limitation(s)
		and adverse childhood experiences had changes in the cingulate gyrus, inferior parietal lobule, and precuneus	
Khan et al,[58] 2020	Evaluate studies exploring the covariation among pain, depression, and/or anxiety using a twin study design	Six studies examined the covariation between pain and depression. One classic twin study reported shared genetic influences entirely explained the association, three reported significant nonshared environmental effects, one co-twin control study reported unmeasured genetic and shared environmental effects explained the association, and 1 reported no evidence for unmeasured familial factors	Five of the six studies were cross-sectional
Dresler et al,[59] 2019	Explore the association between migraine and psychiatric disorders (including depression)	A total of 178 studies included. For depression, potential neurobiologic mechanisms include shared genes, neurotransmitter systems (serotonin, dopamine, and GABA), shared HPA axis involvement, cytokine imbalances, and neuroplasticity. From a psychologic perspective, stress and personality traits (mostly neuroticism) have been implicated	Most studies were cross-sectional

* ALE, activation likelihood estimation; GABA, γ-aminobutyric acid; HPA, hypothalamic pituitary axis; MRI, magnetic resonance imaging; VBM, voxel-based morphometry.

examined adverse childhood experience and major depressive disorder and chronic pain in combination). Most (40 studies) were considered at low risk of bias. Task-based fMRI results indicated that (compared with people without adverse childhood experiences) those with depression and adverse childhood experiences had functional and structural brain changes in the superior frontal, lingual gyrus, hippocampus, insula, putamen, superior temporal, inferior temporal gyrus, and anterior cerebellum, and those with depression or chronic pain and adverse childhood experiences had changes in the cingulate gyrus, inferior parietal lobule, and precuneus. Although this suggests adverse childhood experiences can lead to brain alterations, without examining changes in people with both chronic pain and depression, their relevance to this relationship is unknown.[57]

Second, Khan et al narratively synthesized findings from twin studies exploring the covariation between pain, depression, and/or anxiety. From 23 included studies, six examined the covariation between pain and depression. Marked heterogeneity was seen in study methods (eg, depression definitions), and all but one were cross-sectional. Results varied, with one study reporting that the association between pain and depression was "entirely explained" by shared genetic influences, three reporting significant nonshared environmental effects underlying the association, one reporting the association was explained by unmeasured genetic and shared environmental effects, and one reporting no evidence of unmeasured familial factors.[58] Consequently, the extent to which genetic and environmental factors underlie the association is unclear.

Third, Dresler et al explored the association between migraine and psychiatric disorders (including depression) and its potential mechanisms. For major depression, they reported that studies indicated a bidirectional relationship, with potential mechanisms including shared genes, neurotransmitter changes, shared hypothalamic pituitary adrenal axis involvement, cytokine imbalances, and neuroplasticity.[59]

Summary. Evidence for causal pathways for the association between pain and depression in IA is limited, with only one longitudinal study showing evidence for a potential causal pathway (suggesting passive pain-coping behaviors increase the risk of depression). In non-IA populations, many studies have examined the potential role of neuroplasticity, with various structural/functional brain changes observed in people with pain and depression; however, whether these changes are clinically relevant remains unknown. In non-IA populations, the clearest evidence relates to the role of sleep disturbance, which is estimated to account for 12.5% of the effect of depression on pain.

Clinical assessment of pain and depression in IA

Overview. *This section details evidence on the potential ways to assess pain and depression in adults with IA in routine care. It summarises recommendations from clinical guidelines and systematic review evidence on outcome measures*.

Guidelines on pain assessments. Most IA guidelines focus on disease activity, providing little guidance for pain assessment.[62–66] Only one guideline (from EULAR) has specifically focused on IA pain. This emphasizes the need for biopsychosocial pain assessments, considering (1) patients' views on their pain's cause; (2) pain severity (using a VAS/NRS) and characteristics; (3) current/previous pain treatments; (4) current inflammation/joint damage as pain sources; and (5) pain-related biologic, psychologic, and social factors.[67] Undertaking these in a brief outpatient visit/primary care consultation is challenging. A UK guideline, which will cover pain assessments in patients with IA, is currently being produced by the British Society for Rheumatology.[68]

Guidelines on depression assessments. Although no specific guideline exists for mental health assessments in IA, the EULAR and National Institute For Health and Care Excellence (NICE) IA guidelines highlight the importance of assessing mental health.[63,69] The NICE RA guideline directs people to NICE "Depression in adults with a chronic physical health problem" guidance, which proposes using a simplified Patient Health Questionnaire (PHQ) 2 version for case identification, asking people who may have depression two questions: (1) "during the last month, have you often been bothered by feeling down, depressed, or hopeless?" and (2) "during the last month, have you often been bothered by having little interest or pleasure in doing things?"[70] If people answer "yes" to either, it recommends further exploration of symptoms by a practitioner competent in assessing mental health. Within the United States, the Preventive Services Task Force produced a recommendation statement in 2023 on screening for depression in adults. This recommends screening adults for depression in primary care. Although no single approach is advocated, the statement highlights the availability of commonly used screening instruments, including the PHQ, CES-D, and (in older adults) the Geriatric Depression Scale. It acknowledges the lack of evidence for optimal timing for screening, with a pragmatic approach suggested of screening adults that have not been screened before and using clinical judgment to determine if additional screening of patients at increased risk is warranted.[71]

Tools to assess pain and depression in IA. We identified 19 systematic reviews evaluating pain and/or depression outcome measures in IA (from 867 references). We have described findings from five key systematic reviews of most clinical relevance (with many considering PROMs in research settings). Full details of all reviews are provided in Table 3. First, Englbrecht et al reviewed literature on measuring pain in IA to generate clinical practice recommendations. Across 51 studies, single pain-

Table 3. Systematic reviews examining outcome measures for pain and depression in IA

Study, year	Aim	Key findings	Limitation(s)
Pain outcome measures			
Rutter-Locher et al,[45] 2023	Evaluate self-reported pain sensitivity and neuropathic-like pain in IA	Sixty-three studies included. The self-reported questionnaires used were the CSI (11 studies), DN4 (17 studies), LANSS (seven studies), painDETECT (33 studies), and McGill (SF-McGill [three studies] and McGill [one study]). Pain sensitivity and neuropathic-like pain in IA reported as having prevalence of 31% to 42%, questionnaire dependent	Significant heterogeneity present
Barnabe et al,[72] 2022	Identify and rate evidence in validation studies for PROMs in populations at risk of inequity	Considered pain VAS (100 mm or 5-point scale) and PROMIS pain measures. Three studies considered pain VAS (all 100 mm) and appeared reliable and feasible across languages with strong construct validity linking pain to markers of inflammation (weak correlations with other disease metrics). One study evaluated PROMIS pain measure and showed cross-cultural adaptability, feasibility, and validity when correlated with established measures	Small number of studies with limited country representation
Ortega-Avila et al,[75] 2019	Identify self-reported outcome measures specific to the foot and ankle in RA and investigate their psychometric properties	Fourteen studies included. Four pain measures identified: Ankle Osteoarthritis Scale, Manchester Foot Pain Disability Index, Rowan Foot Pain Assessment Questionnaire, and Self-Reported Foot and Ankle Score. Self-Reported Foot and Ankle Score had the best overall psychometric properties	Incomplete data from some studies
Englbrecht et al,[73] 2012	Review literature on measuring pain (and its treatment) in IA to generate clinical practice recommendations	Fifty-one articles included. Data showed that single pain-related items such as the VAS, NRS, or verbal rating scale provide sufficient clinimetric information and can be recommended as an overall pain measure in clinical practice. Similar results obtained for pain subscales of the AIMS/AIMS2 and bodily pain subscale of SF-36	Search term limited to "pain measures"; may have overlooked some pain outcome measures
van der Leeden et al,[76] 2008	Create an instrument inventory of measures for foot function, pain, and related disability in RA (and clinimetric qualities)	A total of 194 included papers. Six feet pain instruments identified including one-item questions (eg, pain VAS) and a questionnaire (BPI). Nineteen studies and 16 instruments identified for clinimetric evaluation. Six measured both foot pain and disability (two scoring systems; four questionnaires) with varying clinimetric properties and most demonstrating gaps in validation	Search specific to current instruments; may have overlooked older outcome measures
Depression outcome measures			
Victoria et al,[74] 2023	Determine most accurate depression scale in RA	Twenty-eight studies included. Eleven tools evaluated, of which the CES-D (and its variations) were used most frequently and considered to have the best psychometric properties, including internal consistency, sensitivity, and specificity	Methodologic heterogeneity across studies
Dickens et al,[77] 2002	Examine the association between RA and depression (including the influence of the depression assessment method)	Twelve studies included. Ten used a single measure of depression; three used HADS (other measures include GCS, PDI, IDD, POMS, CES-D, AIMS, and Kasl D). The effect sizes obtained from studies using HADS were different from those using other methods, with studies using HADS finding patients with RA to be more depressed than studies using other measures	Historical review
Both pain and depression outcome measures			
Teuwen et al,[78] 2023	Describe the use of PROMIS measures in clinical studies involving people with RA or axial SpA	Twenty-nine studies included. PROMIS Pain Interference (17 studies) consistently indicated poor health outcomes and was employed in various formats (eg, item banks, computer adaptive tests, and short forms) highlighting its adaptability across study designs. PROMIS Depression (12 studies) similarly was applied in various forms. A greater	Psychometric properties of PROMIS measures were not reviewed

(Continued)

Table 3. (*Cont'd*)

Study, year	Aim	Key findings	Limitation(s)
		standardization in the use of PROMIS measures in clinical studies is required	
Hansen et al,[79] 2022	Identify the use of outcome domains and outcome measures across all self-management trials in IA	Thirty-eight studies included. Twenty-five trials considered pain; pain measures included VAS (16), NRS (6), AIMS2 (2), and HAQ (1). A total of 64% of trials measuring pain used VAS. Sixteen trials considered depression/anxiety. Depression measures included HADS (9), CES-D (6), AIMS (2), DS (1), and the Zung depression scale (1). A total of 47% of trials measuring depression/anxiety used HADS	Excluded nonvalidated measures
Küçükdeveci et al,[80] 2021	Describe RA PROMs used over the past 20 years and their performance metrics	A total of 496 articles were included and 125 PROMs identified. Twenty pain PROMs identified: 4 had moderate psychometric evidence (AIMS, AIMS-2, Nottingham Health Profile, and SF-36), 10 weak, and 6 absent. Four PROMs were depression specific; only one had moderate psychometric evidence (HADS depression subscale)	Commonly used outcome measures (ie, pain VAS and PHQ-9) not included
Minnock et al,[81] 2018	Identify patient outcomes measured in RA studies that reported nursing interventions	Seven RCTs and three observational studies included. Fifty-nine measurement instruments were identified, including four pain measures (5-point scale, VAS, AIMS, and RAID). Four depression scales identified (AIMS, NHP, SF-12 MCS, and HADS)	Qualitative studies were excluded
Tang et al,[82] 2017	Review literature on RA PROM data and identify measures implemented in Japan	A total of 100 articles included. Of 23 studies measuring pain, 21 used the VAS. Other validated PROMs were not identified for pain. Eight PROMs were used for depression. The Zung SDS was the most frequently reported, followed by the BDI-2 and HAM-D	Search only conducted in two databases (PubMed and Ichushi Web)
Kilic et al,[83] 2016	Assess frequency of PROM use in RA studies (2013–2015) and compare results with a previous systematic review (2005–2007)	A total of 250 articles included. Pain was reported in 40.0% of studies, most of these used VAS/NRS (89.0%). Pain reporting reduced from previous review (previously 55.9%). Psychologic status reported in 9.6% of studies. HADS (25%) and BDI (25%) most used. There was an increase in the use of depression measures from previous review (previously reported at 7.3%)	Only used one database (PubMed) spanning 2 years

*AIMS, Arthritis Impact Measurement Scale; BDI, Becks Depression Inventory; BPI, Bodily Pain Index; CES-D, Center for Epidemiological Studies Depression; CSI, Central Sensitization Inventory; DN4, Douleur Neuropathique 4; DS, Depression Scale; GCS, Glasgow Depression Scale; HADS-D, Hospital Anxiety Depression Scale-Depression subscale; HAM-D, Hamilton Depression Rating Scale; HAQ, Health Assessment Questionnaire; IA, inflammatory arthritis; IDD, Inventory to Diagnose Depression; LANSS, Leeds Assessment of Neuropathic Symptoms and Signs; NHP, Nottingham Health Profile; NRS, Numeric Rating Scale; PDI, Perinatal Depression Inventory; PHQ-9, Patient Health Questionnaire-9; POMS, Profile of Mood States; PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcomes Measurement Information System; RA, rheumatoid arthritis; RAID, Rapid Assessment, Interface, and Discharge; SDS, Self-Rating Depression Scale; SF, Short Form; SF-12 MCS, Short Form 12 Mental Component Score; SF-36, Short Form 36; SpA, spondyloarthropathy; VAS, Visual Analog Scale.

related items, such as the VAS or NRS, were considered well validated and suitable for routine clinical use owing to strong clinimetric properties (convergent validity, responsiveness, and feasibility).[73] Second, Barnabe et al evaluated validation studies for PROMs in populations at risk of inequities; three studies evaluated the pain VAS, which appeared reliable and feasible across languages (Spanish and Japanese).[72]

Third, Rutter-Locher et al evaluated the prevalence of pain sensitivity/neuropathic-like pain in studies using PROMs. A total of 11 studies used the Central Sensitisation Inventory, 17 studies used the Douleur Neuropathique 4 questionnaire, 7 studies used the Leeds Assessment of Neuropathic Symptoms and Signs questionnaire, 33 studies used the painDETECT questionnaire, and 4 studies used the McGill questionnaire. The prevalence of pain sensitivity and neuropathic-like pain ranged from 31% to 42%, highlighting the need to consider the type of pain in assessments.[45]

Fourth, Victoria et al sought to determine the most accurate depression scale in patients with RA. Across the 28 studies, 11 tools were evaluated, of which the CES-D (and its variations) were used most frequently and considered to have the best psychometric properties.[74] Finally, Küçükdeveci et al synthesized evidence of performance (eg, validity, ability to detect meaningful differences, and feasibility) of PROMs in RA to help clinicians make informed choices about which to use for a given purpose. A process to assess the quality of reported psychometric evidence for PROMs (based on the Outcome Measures in RA Clinical Trials filter) was applied. Twenty PROMs to assess pain were identified: four (AIMS, AIMS-2, Nottingham Health Profile, and SF-36) were considered to have moderate, 10 weak, and six absent underpinning psychometric evidence. Of 23 PROMs to assess emotional function/mental health, four were depression specific, and only one was considered to have moderate psychometric evidence (HADS depression subscale). Notably, the widely used pain VAS/NRS and PHQ-9 were not included.[80]

Summary. Current IA guidelines lack detailed recommendations about the assessment of pain and depression and, despite a range of PROMs existing to assess these in patients with IA, the evidence for their psychometric properties is often considered limited. However, it is our perspective that, in view of the dominance of pain in patients with IA, all IA consultations should incorporate the use of PROMs to routinely measure pain (eg, an NRS, which has robust supportive evidence and is quick to complete) and screening for depression where relevant. In our own clinical practice, the use of electronic PROMs to achieve this is highly acceptable to patients.[84]

Future research

Our review has highlighted two research needs. First is a requirement to better understand the causal pathways underpinning the bidirectional relationship between pain and depression in IA. These are likely to include biologic (eg, neuroplasticity), psychologic (eg, pain-coping strategies), and social (eg, work impact) factors. Understanding pain mechanisms is essential to develop novel pain therapies. However, regardless of causality and mechanisms, there is a compelling argument to routinely assess and manage pain and depression in patients with IA, as improving both is likely to have beneficial effects on patients' lives. Therefore, the second research need is to explore how to best integrate pain assessments and depression screening into routine IA care in a manner that is acceptable to patients and clinicians and fair. Although digital systems show promise in achieving this, they have the potential to worsen care inequalities; for instance, older adults (in whom IA is more common[85]) are less likely to have internet access[86] or the skills necessary to navigate online resources independently.[87] Ensuring that groups of people are not left behind in digitally evolving health care systems is crucial.

Conclusions

Many longitudinal studies demonstrate a relationship between pain and depression in people with IA, which appears to be bidirectional. Although the causal pathways underpinning this are unclear, and the effect sizes are small to modest, the prevalent nature of chronic pain and depression in patients with IA supports the role of routinely assessing both issues in outpatient consultations and implementing systemic care approaches to effectively manage them. Further research is needed to understand how to best achieve this in a manner that is feasible and acceptable to patients and clinicians.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published.

REFERENCES

1. Radawski C, Genovese MC, Hauber B, et al. Patient perceptions of unmet medical need in rheumatoid arthritis: a cross-sectional survey in the USA. Rheumatol Ther 2019;6(3):461–471.
2. Taylor P, Manger B, Alvaro-Gracia J, et al. Patient perceptions concerning pain management in the treatment of rheumatoid arthritis. J Int Med Res 2010;38(4):1213–1224.
3. Heiberg T, Kvien TK. Preferences for improved health examined in 1,024 patients with rheumatoid arthritis: pain has highest priority. Arthritis Rheum 2002;47(4):391–397.
4. McWilliams DF, Dawson O, Young A, et al. Discrete trajectories of resolving and persistent pain in people with rheumatoid arthritis despite undergoing treatment for inflammation: results from three UK cohorts. J Pain 2019;20(6):716–727.
5. Pisaniello HL, Lester S, Russell O, et al. Trajectories of self-reported pain-related health outcomes and longitudinal effects on medication use in rheumatoid arthritis: a prospective cohort analysis using the

Australian Rheumatology Association Database (ARAD). RMD Open 2023;9(3):e002962.

6. Strand V, Deodhar A, Alten R, et al. Pain and fatigue in patients with ankylosing spondylitis treated with tumor necrosis factor inhibitors: multinational real-world findings. J Clin Rheumatol 2021;27(8):e446–e455.
7. Conaghan PG, Alten R, Deodhar A, et al. Relationship of pain and fatigue with health-related quality of life and work in patients with psoriatic arthritis on TNFi: results of a multi-national real-world study. RMD Open 2020;6(2):e001240.
8. Hirano F, van der Heijde D, van Gaalen FA, et al. Determinants of the patient global assessment of well-being in early axial spondyloarthritis: 5-year longitudinal data from the DESIR cohort. Rheumatology (Oxford) 2021;60(1):316–321.
9. Euesden J, Matcham F, Hotopf M, et al. The relationship between mental health, disease severity, and genetic risk for depression in early rheumatoid arthritis. Psychosom Med 2017;79(6):638–645.
10. Pollard LC, Choy EH, Gonzalez J, et al. Fatigue in rheumatoid arthritis reflects pain, not disease activity. Rheumatology (Oxford) 2006;45(7):885–889.
11. Scott IC, Daud N, Bailey J, et al. Gabapentinoid use and the risk of fractures in patients with inflammatory arthritis: nested case-control study in the Clinical Practice Research Datalink Aurum. BMC Med 2024;22(1):575.
12. Cox N, Mallen CD, Scott IC. Pharmacological pain management in patients with rheumatoid arthritis: a narrative literature review. BMC Med 2025;23(1):54.
13. Arthritis Research UK. A research roadmap for pain. Accessed November 1, 2025. https://www.versusarthritis.org/media/1672/research-roadmap-pain.pdf
14. Matcham F, Rayner L, Steer S, et al. The prevalence of depression in rheumatoid arthritis: a systematic review and meta-analysis: reply. Rheumatology (Oxford) 2014;53(3):578–579.
15. Vestergaard SB, Esbensen BA, Klausen JM, et al. Prevalence of anxiety and depression and the association with self-management behaviour in >12 000 patients with inflammatory rheumatic disease: a cross-sectional nationwide study. RMD Open 2024;10(1):e003412.
16. World Health Organization. Depressive disorder (depression). Published March 31, 2023. Accessed July 1, 2025. https://www.who.int/news-room/fact-sheets/detail/depression
17. Mattey DL, Dawes PT, Hassell AB, et al. Effect of psychological distress on continuation of anti-tumor necrosis factor therapy in patients with rheumatoid arthritis. J Rheumatol 2010;37(10):2021–2024.
18. Ho RCM, Fu EHY, Chua ANC, et al. Clinical and psychosocial factors associated with depression and anxiety in Singaporean patients with rheumatoid arthritis. Int J Rheum Dis 2011;14(1):37–47.
19. Ang DC, Choi H, Kroenke K, et al. Comorbid depression is an independent risk factor for mortality in patients with rheumatoid arthritis. J Rheumatol 2005;32(6):1013–1019.
20. Kojima M, Kojima T, Suzuki S, et al. Alexithymia, depression, inflammation, and pain in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2014;66(5):679–686.
21. Xu X, Shen B, Zhang A, et al. Anxiety and depression correlate with disease and quality-of-life parameters in Chinese patients with ankylosing spondylitis. Patient Prefer Adherence 2016;10:879–885.
22. Ben Tekaya A, Mahmoud I, Hamdı I, et al. Depression and anxiety in spondyloarthritis: prevalence and relationship with clinical parameters and self-reported outcome measures. Turk Psikiyatri Derg 2019;30(2):90–98.
23. Adnine A, Nadiri K, Soussan I, et al. Mental health problems experienced by patients with rheumatic diseases during COVID-19 pandemic. Curr Rheumatol Rev 2021;17(3):303–311.
24. Swinscow TDV. Correlation and regression. In: Statistics at Square One. 9th ed. BMJ Publishing Group; 1997.
25. LaMorte WW. PH717 Module 9: correlation and regression evaluating association between two continuous variables. Accessed November 1, 2025. https://www.bu.edu/sph/online-mph-and-teaching-public-health/
26. Gialouri CG, Evangelatos G, Zhao SS, et al. Depression and anxiety in a real-world psoriatic arthritis longitudinal study: should we focus more on patients' perception? Clin Exp Rheumatol 2023;41(1):159–165.
27. Chaney JM, Mullins LL, Wagner JL, et al. A longitudinal examination of causal attributions and depression symptomatology in rheumatoid arthritis. Rehabil Psychol 2004;49(2):126–133.
28. Graham-Engeland JE, Zawadzki MJ, Slavish DC, et al. Depressive symptoms and momentary mood predict momentary pain among rheumatoid arthritis patients. Ann Behav Med 2016;50(1):12–23.
29. Ward MM. Are patient self-report measures of arthritis activity confounded by mood? A longitudinal study of patients with rheumatoid arthritis. J Rheumatol 1994;21(6):1046–1050.
30. Snoeck Henkemans SVJ, Vis M, Koc GH, et al. Association between depression and anxiety and inability to achieve remission in rheumatoid arthritis and psoriatic arthritis. Rheumatology (Oxford) Published online November 6, 2024. https://doi.org/10.1093/rheumatology/keae621
31. Rathbun AM, England BR, Mikuls TR, et al. Relationship between depression and disease activity in United States veterans with early rheumatoid arthritis receiving methotrexate. J Rheumatol 2021;48(6):813–820.
32. Wolfe F, Hawley DJ. The relationship between clinical activity and depression in rheumatoid arthritis. J Rheumatol 1993;20(12):2032–2037.
33. McQuillan J, Andersen JA, Berdahl TA, et al. Associations of rheumatoid arthritis and depressive symptoms over time: are there differences by education, race/ethnicity, and gender? Arthritis Care Res (Hoboken) 2022;74(12):2050–2058.
34. Conner TS, Tennen H, Zautra AJ, et al. Coping with rheumatoid arthritis pain in daily life: within-person analyses reveal hidden vulnerability for the formerly depressed. Pain 2006;126(1-3):198–209.
35. Husted JA, Tom BD, Farewell VT, et al. Longitudinal study of the bidirectional association between pain and depressive symptoms in patients with psoriatic arthritis. Arthritis Care Res (Hoboken) 2012;64(5):758–765.
36. Brown GK. A causal analysis of chronic pain and depression. J Abnorm Psychol 1990;99(2):127–137.
37. Smedstad LM, Vaglum P, Moum T, et al. The relationship between psychological distress and traditional clinical variables: a 2 year prospective study of 216 patients with early rheumatoid arthritis. Br J Rheumatol 1997;36(12):1304–1311.
38. Odegård S, Finset A, Mowinckel P, et al. Pain and psychological health status over a 10-year period in patients with recent onset rheumatoid arthritis. Ann Rheum Dis 2007;66(9):1195–1201.
39. Hawley DJ, Wolfe F. Anxiety and depression in patients with rheumatoid arthritis: a prospective study of 400 patients. J Rheumatol 1988;15(6):932–941.
40. International Association for the Study of Pain. IASP Announces Revised Definition of Pain. Published July 16, 2020. Accessed July 1, 2025. https://www.iasp-pain.org/publications/iasp-news/iasp-announces-revised-definition-of-pain/
41. Raja SN, Carr DB, Cohen M, et al. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. Pain 2020;161(9):1976–1982.

42. International Association for the Study of Pain. Terminology. Accessed August 1, 2024. https://www.iasp-pain.org/resources/terminology/?ItemNumber=1698#Pain
43. Al Mohamad F, Rios Rodriguez V, Haibel H, 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.
44. Khot S, Tackley G, Choy E. How to distinguish non-inflammatory from inflammatory pain in RA? Curr Rheumatol Rep 2024;26(12):403–413.
45. Rutter-Locher Z, Arumalla N, Norton S, et al. A systematic review and meta-analysis of questionnaires to screen for pain sensitisation and neuropathic like pain in inflammatory arthritis. Semin Arthritis Rheum 2023;61:152207.
46. Duffield SJ, Miller N, Zhao S, et al. Concomitant fibromyalgia complicating chronic inflammatory arthritis: a systematic review and meta-analysis. Rheumatology (Oxford) 2018;57(8):1453–1460.
47. Freynhagen R, Parada HA, Calderon-Ospina CA, et al. Current understanding of the mixed pain concept: a brief narrative review. Curr Med Res Opin 2019;35(6):1011–1018.
48. Jesulola E, Micalos P, Baguley IJ. Understanding the pathophysiology of depression: from monoamines to the neurogenesis hypothesis model - are we there yet? Behav Brain Res 2018;341:79–90.
49. Ionescu CE, Popescu CC, Agache M, et al. Depression in rheumatoid arthritis: a narrative review-diagnostic challenges, pathogenic mechanisms and effects. Medicina (Kaunas) 2022;58(11):1637.
50. Brown GK, Nicassio PM, Wallston KA. Pain coping strategies and depression in rheumatoid arthritis. J Consult Clin Psychol 1989;57(5):652–657.
51. Chen MH, Sun CK, Lin IM, et al. Size reduction of the right amygdala in chronic pain patients with emotional stress: a systematic review and meta-analysis. Pain Med 2023;24(5):556–565.
52. Zheng CJ, Van Drunen S, Egorova-Brumley N. Neural correlates of co-occurring pain and depression: an activation-likelihood estimation (ALE) meta-analysis and systematic review. Transl Psychiatry 2022;12(1):196.
53. Ma T, Ji YY, Yan LF, et al. Gray matter volume abnormality in chronic pain patients with depressive symptoms: a systemic review and meta-analysis of voxel-based morphometry studies. Front Neurosci 2022;16:826759.
54. Malfliet A, Coppieters I, Van Wilgen P, et al. Brain changes associated with cognitive and emotional factors in chronic pain: a systematic review. Eur J Pain 2017;21(5):769–786.
55. Thompson T, Correll CU, Gallop K, et al. Is pain perception altered in people with depression? A systematic review and meta-analysis of experimental pain research. J Pain 2016;17(12):1257–1272.
56. Karimi R, Mallah N, Scherer R, et al. Sleep quality as a mediator of the relation between depression and chronic pain: a systematic review and meta-analysis. Br J Anaesth 2023;130(6):747–762.
57. Antoniou G, Lambourg E, Steele JD, et al. The effect of adverse childhood experiences on chronic pain and major depression in adulthood: a systematic review and meta-analysis. Br J Anaesth 2023;130(6):729–746.
58. Khan WU, Michelini G, Battaglia M. Twin studies of the covariation of pain with depression and anxiety: a systematic review and re-evaluation of critical needs. Neurosci Biobehav Rev 2020;111:135–148.
59. Dresler T, Caratozzolo S, Guldolf K, et al; European Headache Federation School of Advanced Studies (EHF-SAS). Understanding the nature of psychiatric comorbidity in migraine: a systematic review focused on interactions and treatment implications. J Headache Pain 2019;20(1):51.
60. Martucci KT, Ng P, Mackey S. Neuroimaging chronic pain: what have we learned and where are we going? Future Neurol 2014;9(6):615–626.
61. Kirby LAJ, Robinson JL. Affective mapping: an activation likelihood estimation (ALE) meta-analysis. Brain Cogn 2017;118:137–148.
62. Fraenkel L, Bathon JM, England BR, et al. 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2021;73(7):1108–1123.
63. National Institute for Health and Care Excellence. Rheumatoid arthritis in adults: management. Updated October 12, 2020. Accessed November 11, 2025. https://www.nice.org.uk/guidance/ng100
64. National Institute for Health and Care Excellence. Spondyloarthritis in over 16s: diagnosis and management. Updated June 2, 2017. Accessed November 1, 2025. https://www.nice.org.uk/guidance/ng65
65. Singh JA, Guyatt G, Ogdie A, et al. Special article: 2018 American College of Rheumatology/National Psoriasis Foundation guideline for the treatment of psoriatic arthritis. Arthritis Rheumatol 2019;71(1):5–32.
66. Ward MM, Deodhar A, Gensler LS, et al. 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. Arthritis Rheumatol 2019;71(10):1599–1613.
67. 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(6):797–807.
68. Scott IC, Babatunde O, Barker C, et al. Pain management in people with inflammatory arthritis: British Society for Rheumatology guideline scope. Rheumatol Adv Pract 2024;8(4):rkae128.
69. 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(10):1278–1285.
70. National Institute for Health and Care Excellence. Depression in adults with a chronic physical health problem: recognition and management. Published October 28, 2009. Accessed November 1, 2025. https://www.nice.org.uk/guidance/cg91
71. Barry MJ, Nicholson WK, Silverstein M, et al; US Preventive Services Task Force. Screening for depression and suicide risk in adults: US Preventive Services Task Force recommendation statement. JAMA 2023;329(23):2057–2067.
72. Barnabe C, Wattiaux A, Petkovic J, et al. Validation studies of rheumatoid arthritis patient-reported outcome measures in populations at risk for inequity: a systematic review and analysis using the OMERACT summary of measurement properties equity table. Semin Arthritis Rheum 2022;55:152029.
73. Englbrecht M, Tarner IH, van der Heijde DM, et al. Measuring pain and efficacy of pain treatment in inflammatory arthritis: a systematic literature review. J Rheumatol Suppl 2012;90:3–10.
74. Victoria MA, Lucio VR, Cristina HD. Patient self-reported instruments for assessing symptoms in rheumatoid arthritis. Rheumatol Int 2023;43(10):1781–1790.
75. Ortega-Avila AB, Ramos-Petersen L, Cervera-Garvi P, et al. Systematic review of the psychometric properties of patient-reported outcome measures for rheumatoid arthritis in the foot and ankle. Clin Rehabil 2019;33(11):1788–1799.
76. van der Leeden M, Steultjens MPM, Terwee CB, et al. A systematic review of instruments measuring foot function, foot pain, and foot-related disability in patients with rheumatoid arthritis. Arthritis Rheum 2008;59(9):1257–1269.

77. Dickens C, McGowan L, Clark-Carter D, et al. Depression in rheumatoid arthritis: a systematic review of the literature with meta-analysis. Psychosom Med 2002;64(1):52–60.
78. Teuwen MMH, Knaapen IRE, Vliet Vlieland TPM, et al. The use of PROMIS measures in clinical studies in patients with inflammatory arthritis: a systematic review. Qual Life Res 2023;32(10):2731–2749.
79. Hansen CW, Esbensen BA, de Thurah A, et al. Outcome measures in rheumatology applied in self-management interventions targeting people with inflammatory arthritis: a systematic review of outcome domains and measurement instruments. Semin Arthritis Rheum 2022;54:151995.
80. Küçükdeveci AA, Elhan AH, Erdoğan BD, et al. Use and detailed metric properties of patient-reported outcome measures for rheumatoid arthritis: a systematic review covering two decades. RMD Open 2021;7(2):e001707.
81. Minnock P, McKee G, Kelly A, et al. Nursing sensitive outcomes in patients with rheumatoid arthritis: a systematic literature review. Int J Nurs Stud 2018;77:115–129.
82. Tang AC, Kim H, Crawford B, et al. The use of patient reported outcome measures for rheumatoid arthritis in Japan: a systematic literature review. Open Rheumatol J 2017;11(1):43–52.
83. Kilic L, Erden A, Bingham CO III, et al. The reporting of patient-reported outcomes in studies of patients with rheumatoid arthritis: a systematic review of 250 articles. J Rheumatol 2016;43(7):1300–1305.
84. Cox N, Kettle C, Wang H, et al. E041 Feasibility and acceptability of electronic patient reported outcome measures in the routine care of people with inflammatory arthritis: Haywood Arthritis Portal study. Rheumatology 2024;63(Supplement_1):keae163.269.
85. Scott IC, Whittle R, Bailey J, et al. Rheumatoid arthritis, psoriatic arthritis, and axial spondyloarthritis epidemiology in England from 2004 to 2020: an observational study using primary care electronic health record data. Lancet Reg Health Eur 2022;23:100519.
86. Office for National Statistics. Internet access - households and individuals QMI. Updated August 8, 2019. Accessed November 1, 2025. https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/methodologies/internetaccesshouseholdsandindividualsqmi
87. Lloyds Bank. UK Consumer Digital Index. Accessed November 1, 2025. https://www.lloydsbank.com/banking-with-us/whats-happening/consumer-digital-index.html

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 31–37
DOI 10.1002/acr.25636

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

REVIEW ARTICLE

Considerations for Issues of Regression to the Mean and Contextual Effects in Clinical Trials for Pain in Rheumatic Diseases

Yen T. Chen, Guohao Zhu, Afton L. Hassett, Daniel Clauw, and Susan L. Murphy

Recently, there has been growing discussion about how to best assess pain in clinical trials in rheumatic diseases. Reliable measurement of pain outcomes is essential for accurately determining the effectiveness of treatments. Although pain intensity is the most common measure of change in pain trials, other pain-related measures, such as pain interference, are also frequently assessed. Interpreting treatment effects on these outcomes can be complicated due to statistical phenomena, particularly regression to the mean and contextual effects. These issues can substantially distort clinical trial findings, potentially leading to inaccurate conclusions about the efficacy of interventions. The present article provides an overview of regression to the mean and contextual effects, emphasizing their implications for internal validity, clinical decision-making, and ethical considerations in trials. Additionally, this article highlights key study design and analysis considerations, including methodologic and statistical approaches, that researchers can implement to mitigate or better account for these challenges. Practical recommendations are offered to enhance the rigor of pain assessment, with specific attention to osteoarthritis as a representative example within rheumatic disease research. By recognizing and addressing regression to the mean and contextual effects proactively, researchers can strengthen trial outcomes, improve clinical interpretations, and support the identification of effective treatments.

Introduction

In clinical trials designed to examine treatments for people with rheumatic diseases, pain improvement is often a key objective. However, there are difficulties in designing and interpreting the results of clinical trials when pain is the study outcome. Pain is complex and varies across individuals by intensity, persistence, location, quality, and mechanism (eg, nociceptive, neuropathic, nociplastic). Whereas reviews have discussed considerations for selecting comprehensive pain assessments, including psychosocial factors and related impact on functioning,[1,2] this article focuses on methodologic concerns that affect pain assessment in clinical trials. In particular, we highlight two critical issues, regression to the mean and contextual effects, both of which can significantly influence trial outcomes. Although these issues are relevant for trials measuring pain intensity, similar considerations may also apply to other pain-related outcomes, such as pain interference. We describe how regression to the mean and contextual effects arise and outline strategies to better understand and mitigate their impact to improve the interpretation and validity of clinical trial findings.

Regression to the mean

Regression to the mean refers to the phenomenon in which extreme measurements tend to be closer to the average when measured in subsequent time points.[3] Among people with chronic pain, such as osteoarthritis (OA), these extreme baseline ratings may reflect temporary pain exacerbation or natural day-to-day fluctuations.[3,4] As a result, even without any treatment, follow-up pain ratings may appear improved simply because the initial measurement captured an unusually high level of pain. Researchers may be incorporating this bias in trials because they often recruit people with moderate to high pain in their studies so that there is room for participants to improve with treatment. Additionally, potential participants may be more apt to seek out studies when they are experiencing higher levels of pain to get some relief. Having participants who have higher pain than the population of people with OA may be contributing to regression

Yen T. Chen, PhD, Guohao Zhu, PhD, Afton L. Hassett, PsyD, Daniel Clauw, MD, Susan L. Murphy, ScD: University of Michigan, Ann Arbor.
Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25636.

Address correspondence via email to Yen T. Chen, PhD, at yentchen@med.umich.edu.
Submitted for publication August 6, 2025; accepted August 7, 2025.

to the mean. Current research guidelines suggest including participants who have a pain rating of at least mild to moderate pain (4 of 10).[5,6] Although it is important for individuals to have some level of pain to be appropriate for studies investigating pain improvement, the higher the pain threshold for the inclusion criterion, the more the results will be influenced by regression to the mean.[7] If most participants in a sample have elevated pain ratings compared to their usual pain, there will be an effect of regression to the mean that is not a true treatment effect.[8] Treatment effects may especially be overestimated due to regression to the mean for within-group comparisons or in one-arm trials. This biased treatment effect is due to the natural improvement in a placebo arm that is not accounted for when only one group's data are being analyzed.[3] In these cases, the minimal clinically important difference (MCID) should be interpreted cautiously.[3]

There are several strategies that can help reduce the effect of regression to the mean (see Table 1). These strategies can be implemented during both the design and analysis phases to improve the validity and interpretability of study findings.

During the designing phase, researchers can consider incorporating a run-in period with multiple baseline measurements to reduce variability due to short-term fluctuations in pain.[9] This approach helps to ensure more stable baseline estimates and to minimize the likelihood that participants are enrolled during a temporary pain flare, which could otherwise result in exaggerated improvements due to regression to the mean. In addition to a run-in period, researchers can consider including a coprimary outcome measure that is less susceptible to short-term changes.[10] Using multiple outcome metrics, such as combining pain intensity ratings with patient global impression of change[11] or physical function,[12] provides a more robust assessment of treatment effects and can help verify whether observed changes are consistent or are likely driven by regression to the mean. Regression to the mean seems to be most pronounced for pain intensity, particularly when participants are selected based on moderate to high pain levels, and evidence suggests that other patient-reported outcomes such as physical function and quality of life may be less susceptible to this effect. For example, Englund and Turkiewicz found that the degree of regression to the mean was relatively small for physical function and quality of life compared to knee pain intensity, likely because trial enrollment typically conditions on pain rather than function or quality-of-life scores.[13] Incorporating a broader range of outcomes may therefore help reduce overinterpretation of changes driven by statistical artifacts. Researchers may also consider incorporating a control group, which is one of the most effective ways to address regression to the mean. By randomizing participants into both intervention and control arms, the effects of regression are distributed across groups, making it easier to isolate true treatment effects.[8] For studies that use a threshold-based enrollment criterion, such as including only participants with moderate to high baseline pain levels, researchers can consider a regression-discontinuity design.[14] This design compares outcomes for participants just above and below the eligibility cutoff, helping to distinguish intervention effects from regression to the mean.

Design strategies that account for individual variability and extreme baseline scores are also important. Researchers can consider using a crossover design,[15] in which each participant

Table 1. Considerations for RTM in studies and potential strategies*

Design/analysis	Potential issues	Strategies
Design phase		
Moderate to high pain required for study enrollment	Inflates RTM effects	Use run-in periods at baseline to reduce measurement variance/error[9] Including other coprimary outcome measures[10] (eg, patient global impression of change[11] or physical function[12])
Conducted a one-arm trial	Overestimates treatment effects	Conduct a randomized trial with a control group to examine natural course of pain over time[8]
Threshold-based enrollment	RTM amplified at cutoff points	Use a regression-discontinuity design to compare outcomes for participants just above/below cutoff[14]
Inability to control individual variability	RTM distorts within-subject changes	Use crossover designs so participants serve as their own controls[15] Consider stratified controls to balance baseline pain across arms[16]
Analysis phase		
No analyses to examine RTM	Misinterprets study findings	Plot change scores to visually examine the effect of RTM[17] Use ANCOVA to control for individual baseline measure[8]
Limited control for extreme baselines	Exaggerated treatment effects	Apply statistical correction formulas using test-retest correlations[18] Use Bayesian hierarchical modeling to shrink extreme baselines[19,20,21]

* ANCOVA, analysis of covariance; RTM, regression to the mean.

receives both the intervention and control conditions in separate study periods, with a washout period in between. This within-subject design allows participants to serve as their own controls, eliminating between-subject variability and minimizing the influence of regression to the mean, which is expected to occur similarly across both conditions. However, if a participant begins the first study period during a pain flare, the change observed during that period may be more pronounced than that in the second period. This may introduce bias into treatment comparisons. To address this, researchers may consider including repeated pretreatment assessments to better characterize initial symptom levels. Still, crossover designs may not always be feasible, particularly in studies of chronic conditions in which the intervention has a lasting effect or washout periods are impractical. In such cases, researchers can consider using stratified randomization.[16] Stratification involves grouping participants into subgroups (strata) based on baseline variables before randomization, ensuring that the distribution of baseline scores is balanced across study arms. In addition to improving balance at baseline, stratification also allows for data to be analyzed within each stratum and then combined to estimate an overall treatment effect. However, this approach requires a sufficient sample size within each stratum to ensure more accurate estimates and maintain statistical power, which may be challenging in studies with small samples. When feasible, stratified randomization helps control for baseline imbalances and reduce the distortion of treatment effects that can result from regression to the mean.

In the analysis phase, several statistical approaches can be used to detect or adjust for the effects of regression to the mean. A simple yet informative method is to examine individual or group-level change scores by plotting pre- and postintervention data.[17] Visual inspection of scatter plots or line graphs displaying baseline versus follow-up pain intensity ratings can help identify whether improvements are more likely due to natural fluctuations rather than true treatment effects, particularly when participants begin with extreme baseline values. Including these plots provides readers with tangible examples to facilitate understanding and interpretation of potential regression to the mean. A more formal method is to use analysis of covariance (ANCOVA), which includes baseline values as covariates in the model.[8] This approach statistically adjusts for initial score differences between groups and provides a more accurate estimate of the treatment effect by controlling for baseline variability. ANCOVA is especially useful in randomized controlled trials in which some residual imbalance in baseline values may still occur despite randomization. Researchers may also consider applying statistical correction formulas that adjust for the test-retest reliability of the outcome measure.[18] These formulas estimate the amount of change that could be expected due to regression to the mean alone, based on the stability of the measure over time. Incorporating such corrections can help differentiate treatment effects from statistical artifacts and improve the interpretability of study results.

Finally, to address the influence of extreme baseline values more robustly, researchers can consider using Bayesian hierarchical modeling.[19] This analytic technique models individual-level outcomes as drawn from a broader population distribution and applies a shrinkage effect, pulling extreme values toward the group mean. By reducing the weight of outlier scores, especially those that may be disproportionately affected by regression to the mean, Bayesian hierarchical models provide more stable and reliable estimates of treatment effects. This method is particularly advantageous in studies with small sample sizes or high variability at baseline, in which the risk of regression effects is greater.[20,21]

Contextual effects

Beyond the direct effects of the intervention itself, patient outcomes are profoundly influenced by contextual effects, nonspecific elements inherent in the clinical or research setting.[22–25] These contextual effects, which comprise patient and clinician characteristics, along with psychosocial and environmental factors, can alter the experiences of individuals with OA beyond specific treatment.[26] They arise from multiple dimensions of the therapeutic encounter, including patient beliefs and expectations, clinician behaviors (eg empathy, communication style),[27] patient–therapist relationship and/or interactions (eg, trust, nonverbal cues),[28] and the treatment setting.[29] For example, patients may report lower pain scores not because of physiologic improvement but because they trust their clinician, feel supported, or interact with clinicians who express optimism about treatment. Research shows these contextual factors, such as provider's warmth, empathy, and communication style, can significantly influence reported pain outcomes in OA.[30,31] Collectively, these contextual factors can substantially influence treatment outcomes.[26,32–34]

These contextual factors are also important to consider because they trigger psycho-neurobiologic responses within the patient,[25,27] actively modulating pain perception and leading to changes at neurobiologic, perceptual, and cognitive levels.[25] Notably, placebo responses can sometimes exceed the effect size of active treatment, highlighting the substantial influence of contextual effects.[24] Furthermore, contextual effects are not uniformly transient or predictable. Previtali et al found that placebo responses to injections in OA could persist long term and vary across outcome measures, challenging the assumption that these effects are short lived or homogeneous.[35] Positive contexts (eg, supportive clinician interactions) can enhance placebo effects, improving pain relief, whereas negative contexts (eg, distrust or pessimism) may induce nocebo effects, worsening symptoms.[25]

Understanding how contextual effects impact pain perception is crucial for accurately interpreting interventions and optimizing clinical care in OA and other rheumatic diseases.[25,36,37] In randomized controlled trials (RCTs), the primary goal has traditionally been to isolate and quantify the treatment efficacy. This is typically done by comparing the outcomes in an active

intervention group to those in a placebo or sham control group. The placebo arm is intended to capture nonspecific effects, allowing researchers to subtract them and determine the effect solely attributable to the active intervention. However, recent research has highlighted that contextual effects present in both the active treatment arm and placebo or sham arm contribute significantly to the observed outcomes in RCTs.[38] Studies on nonpharmacological interventions, such as exercise and acupuncture, indicate that contextual effects, including placebo effects and natural symptom fluctuations, play a dominant role in perceived pain relief in patients with knee OA.[26] A meta-analysis of OA RCTs found that, on average, 75% of the observed pain reduction was attributable to the contextual effects.[39] The proportion attributable to contextual effect (PCE) can vary depending on the specific treatment. For instance, in OA pain reduction, PCE ranged from 47% for intra-articular glucocorticoids to 91% for joint lavage.[39]

The high proportion of benefit attributable to context gives rise to what is known as the "efficacy paradox." Although RCTs are designed to isolate and test specific treatment effects (which are often modest or below clinically meaningful thresholds), real-world clinical outcomes reflect the combined impact of specific and contextual effects.[40] Because contextual effects are always present in clinical encounters and can contribute substantially to the overall response,[32] a treatment with a small specific effect might still be perceived as clinically beneficial by patients and clinicians.[39,41] For example, Saueressig et al note that the estimated contextual effect for pain intensity is 5.3 on a 0- to 100-point visual analog scale, which makes up more than 25% of the typical MCID for pain (20 points) and 30% to 40% of the effect of commonly recommended treatments such as exercise, psychosocial interventions, and multidisciplinary management.[42] This highlights that the context is a sizable component that should not be disregarded.

The efficacy paradox complicates the interpretation of RCT results. A statistically significant difference between an active treatment and a placebo may reflect only a small specific effect, overshadowed by a large contextual effect present in both groups. Conversely, the absence of a significant difference does not necessarily mean that the total treatment package (comprising both the specific and contextual effects) is ineffective in clinical practice. Thus, focusing solely on the specific effect measured by the difference between active and placebo arms can obscure the substantial benefits that patients may derive from the contextual aspects of care.[39,40] Although enhancing contextual factors has been proposed as a cost-effective strategy to improve outcomes, one meta-analysis suggested that intentionally boosting contextual effects alone in nonpharmacological treatments may offer limited additional clinical benefit.[42] Nevertheless, this does not diminish the essential role contextual factors play in standard care. Contextual effects should be viewed as an integral component of treatment outcomes, not merely as confounding elements to be minimized or subtracted.[43]

To better interpret treatment effects in pain studies for OA and other rheumatic diseases, researchers must actively account for contextual effects. If unaddressed, contextual effects can bias estimates of treatment efficacy and lead to misleading conclusions. A comprehensive strategy to reduce and interpret contextual effects should involve both study design and measurement and analytic approach (see Table 2).

Table 2. Considerations for contextual effects and potential strategies*

Design/analysis	Potential issues	Strategies
Design phase		
One-arm or two-arm RCT (active vs placebo)	Difficult to isolate contextual effects from treatment effects	Use three-arm RCT to add untreated control to quantify contextual effects (placebo vs untreated)[44] Use cluster randomization to control for provider-level biases[45]
Unstandardized patient interactions	Variability in placebo/nocebo responses due to uncontrolled expectations	Use standardized scripts to balance explanations of benefits/risks to minimize bias[22] Include active placebos to mimic side effects[48] Use placebo run-in phase to exclude high responders[49]
Clinician/interventionist variability (eg, empathy, communication)	Placebo effects or nocebo effects from positive or negative interactions	Use blinded outcome assessment (independent evaluators)[50] Have standardized protocol and treatment environment[51]
Inability to disentangle expectancy vs treatment	Biased efficacy estimates	Use balanced placebo designs/open-hidden paradigms[52] Use crossover/randomized withdrawal designs[54]
No qualitative data	Missed insights into patient experiences	Include qualitative data to explain quantitative results[55,56]
Measurement and analysis approach		
Lack of psychosocial measures	Confounding by unmeasured mood/stress/social support	Use expectation rating scales (eg, Credibility/Expectancy Questionnaire)[57] Include validated scales (eg, PROMIS depression, PROMIS anxiety, social support) as covariates/mediators[36,59]
Ignoring proportional contextual effect	Overemphasis on specific effects, undervaluing contextual benefits	Report proportional contextual effect alongside specific/overall effects[39,40]

* PROMIS, Patient-Reported Outcomes Measurement Information System; RCT, randomized controlled trial.

A core design recommendation is the use of a three-arm RCT design that includes an active intervention group, a placebo or sham group, and a nontreated control group.[44] This design allows researchers to estimate the contextual effects by comparing outcomes between the placebo and untreated groups while isolating the specific treatment effect through comparison of the active and placebo groups.[29] This structure also helps adjust for symptom fluctuations and regression to the mean. To further control for provider-related contextual effects, researchers can implement cluster randomization,[45] in which participants are assigned by treatment provider. This approach reduces variability caused by individual differences in clinician behavior or communication. However, when there are a small number of clusters or systematic differences in provider characteristics across study arms (eg, if providers in one arm tend to have better communication skills), cluster randomization may actually increase the risk of biased contextual effects. There are other trade-offs to cluster randomization that have been outlined in several reviews.[45–47]

Additionally, standardizing participant–provider interactions is also critical in reducing contextual effects. Researchers and research personnel should use scripts that provide a consistent explanation of potential treatment benefits and risks. This helps ensure that expectations are managed similarly across study arms. Training health care providers to create a positive therapeutic environment and manage patient expectations can harness the beneficial components of contextual effects while minimizing their confounding influence on the measured efficacy of active treatment.[22] This approach not only improves clinical outcomes but also aligns with a more holistic understanding of pain management.[23] Maintaining blinding of those delivering the intervention is another crucial strategy. The use of active placebos that mimic side effects of the intervention can equalize perceived credibility of treatment arms.[48] Researchers may also consider a placebo run-in phase,[49] which allows for identification and exclusion of individuals who show strong placebo responses before randomization, thereby enhancing the precision of treatment effect estimates.

To reduce bias in outcome measurement, clinical trials should use independent evaluators who are blinded to participants' treatment assignments.[50] Blinding helps make sure that assessments are not influenced by conscious or unconscious expectations about the efficacy of the intervention. This is especially important in trials involving subjective outcomes, such as self-reported pain, in which evaluator interactions or assumptions could unintentionally shape the data collection process. In addition to blinding, maintaining a standardized treatment environment across all study arms is critical.[51] This includes using consistent physical settings, materials, schedules, and procedures so that participants across groups are exposed to similar contextual cues. Standardization helps prevent differences in the care environment from contributing to variability in outcomes, allowing for a more accurate estimation of the treatment's true effect.

For trials specifically aiming to disentangle the effects of patient expectations from the active components of treatment, researchers can use advanced experimental designs such as balanced placebo designs and open-hidden paradigms.[52] These approaches manipulate what participants are told about the treatment to separate expectancy effects from the pharmacological or procedural components. For example, in an open-hidden paradigm, patients might receive an analgesic either openly (with full awareness and explanation from a clinician) or hidden (administered without their knowledge). Research using this design has shown that the same medication can produce stronger pain relief when patients are aware they are receiving it, compared to when it is administered covertly, highlighting the influence of expectations on treatment response. This method has been used in both pharmacological and nonpharmacological pain research to better understand how much of the benefit is attributable to the treatment itself versus the patient's belief in its effectiveness.[33,53] In addition, crossover or randomized withdrawal designs can help evaluate whether observed effects persist or reverse after the intervention is discontinued,[54] providing further insight into the durability and origin of contextual effects.

Finally, because contextual effects often involve subjective experiences, qualitative research is an important complement to quantitative designs. Interviews and open-ended feedback from participants can provide insight into how individuals perceive the intervention and what they believe contributed to their outcomes.[55,56] By incorporating qualitative assessment into quantitative studies or using mixed-methods designs, researchers can better account for the contextual effects of pain, leading to more accurate interpretations of treatment efficacy in pain studies.

Beyond study design, measurement and analytic strategies also play a central role in accounting for contextual effects. One important step is to measure patient expectations at baseline using validated tools such as the Credibility/Expectancy Questionnaire.[57] These scores can later be analyzed to determine whether expectations mediate or moderate treatment outcomes.[58] Other psychosocial variables known to influence pain, such as mood, anxiety, and social support, should also be systematically assessed. Validated instruments such as the Patient-Reported Outcomes Measurement Information System depression and anxiety scales and social support measures can be included as covariates in statistical models.[36,59] Their inclusion allows researchers to adjust for or explore the modifying effects of these psychosocial factors on treatment outcomes. In studies in which the goal is to examine mechanisms, causal mediation analysis can be used to determine whether these contextual variables account for part of the observed treatment effect.[36]

To improve the interpretation of RCT results, researchers are strongly encouraged to report not only the traditional specific treatment effect (difference between active and placebo) but also the overall treatment effect (improvement in the active group) and the PCE.[39,40] This approach provides a more complete,

balanced, and clinically meaningful interpretation. Adhering to reporting guidelines for contextual factors, such as the Control Interventions in Efficacy Trials of Physical, Psychological, and Self-Management Therapies statement,[60] is also advised.[32]

Conclusions

In clinical trials that evaluate pain, two common sources of issues, regression to the mean and contextual effects, can lead to problems in misunderstanding study findings and should be considered in research. This article emphasizes the importance of addressing these issues during both the design and analysis phases of pain research. Strategies such as incorporating three-arm RCTs including run-in periods, collecting data on psychosocial variables, applying appropriate statistical adjustments, and integrating qualitative methods can strengthen the validity of study findings. These strategies provide a more complete understanding of how both specific and nonspecific factors influence outcomes, leading to more accurate and reliable conclusions. By doing so, researchers can design trials that better reflect the complexity of real-world care and contribute to the development of more effective and meaningful interventions for patients.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Edwards RR, Dworkin RH, Turk DC, et al. Patient phenotyping in clinical trials of chronic pain treatments: IMMPACT recommendations. Pain 2016;157(9):1851–1871.
2. Hassett AL, Whibley D, Kratz A, et al. Measures for the assessment of pain in adults. Arthritis Care Res (Hoboken) 2020;72(suppl 10): 342–357.
3. Emanuel KS, Dahmen J, Sierevelt IN, et al. Regression to the mean: statistical bias can mislead interpretation in cartilage and osteoarthritis clinics and research. Cartilage Published online November 11, 2024. https://doi.org/10.1177/19476035241293048
4. Darnall BD. On the importance of baseline pain intensity and measurement methods. Pain 2023;164(9):1887–1888.
5. Dworkin RH, Turk DC, Farrar JT, et al; IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113(1–2):9–19.
6. Langford DJ, Baron R, Edwards RR, et al. What should be the entry pain intensity criteria for chronic pain clinical trials? An IMMPACT update. Pain 2023;164(9):1927–1930.
7. Kamerman PR, Vollert J. Greater baseline pain inclusion criteria in clinical trials increase regression to the mean effect: a modelling study. Pain 2022;163(6):e748–e758.
8. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. Int J Epidemiol 2005;34(1):215–220.
9. Pablos-Méndez A, Barr RG, Shea S. Run-in periods in randomized trials: implications for the application of results in clinical practice. JAMA 1998;279(3):222–225.
10. Hamasaki T, Evans SR, Asakura K. Design, data monitoring, and analysis of clinical trials with co-primary endpoints: a review. J Biopharm Stat 2018;28(1):28–51.
11. Farrar JT, Young JP Jr, LaMoreaux L, et al. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001;94(2):149–158.
12. Patel KV, Allen R, Burke L, et al. Evaluation of composite responder outcomes of pain intensity and physical function in neuropathic pain clinical trials: an ACTTION individual patient data analysis. Pain 2018; 159(11):2245–2254.
13. Englund M, Turkiewicz A. Regression to the mean for physical function and quality of life in clinical trials for symptomatic knee osteoarthritis. Osteoarthritis Cartilage 2025;33(3):391–395.
14. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. J Educ Psychol 1960; 51(6):309–317.
15. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2012;109(15):276–281.
16. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1): 41–55.
17. Linden A. Assessing regression to the mean effects in health care initiatives. BMC Med Res Methodol 2013;13(1):119.
18. Kelly C, Price TD. Correcting for regression to the mean in behavior and ecology. Am Nat 2005;166(6):700–707.
19. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. John Wiley & Sons; 2004.
20. Ahn H, Suchting R, Woods AJ, et al. Bayesian analysis of the effect of transcranial direct current stimulation on experimental pain sensitivity in older adults with knee osteoarthritis: randomized sham-controlled pilot clinical study. J Pain Res 2018;11:2071–2082.
21. Suchting R, Kapoor S, Mathis KB, et al. Changes in experimental pain sensitivity from using home-based remotely supervised transcranial direct current stimulation in older adults with knee osteoarthritis. Pain Med 2020;21(11):2676–2683.
22. Abhishek A, Doherty M. Placebo, nocebo, and contextual effects. In: Doherty M, Hunter DJ, Bijlsma H, Arden N, Dalbeth N, eds. Oxford Textbook of Osteoarthritis and Crystal Arthropathy. Oxford University Press; 2016.
23. Di Blasi Z, Harkness E, Ernst E, et al. Influence of context effects on health outcomes: a systematic review. Lancet 2001;357(9258): 757–762.
24. Doherty M, Dieppe P. The "placebo" response in osteoarthritis and its implications for clinical practice. Osteoarthritis Cartilage 2009;17(10): 1255–1262.
25. Rossettini G, Palese A, Cook C. "Trying to explain the unexplainable": why research on contextual factors in musculoskeletal pain is needed. Pain Manag 2024;14(9):465–468.
26. Chen AT, Shrestha S, Collins JE, et al. Estimating contextual effect in nonpharmacological therapies for pain in knee osteoarthritis: a systematic analytic review. Osteoarthritis Cartilage 2020;28(9):1154–1169.

27. Sherriff B, Clark C, Killingback C, et al. Impact of contextual factors on patient outcomes following conservative low back pain treatment: systematic review. Chiropr Man Therap 2022;30(1):20.
28. Kinney M, Seider J, Beaty AF, et al. The impact of therapeutic alliance in physical therapy for chronic musculoskeletal pain: a systematic review of the literature. Physiother Theory Pract 2020;36(8):886–898.
29. Sandal LF, Thorlund JB, Moore AJ, et al. Room for improvement: a randomised controlled trial with nested qualitative interviews on space, place and treatment delivery. Br J Sports Med 2019;53(6): 359–367.
30. Street RL Jr, Cox V, Kallen MA, et al. Exploring communication pathways to better health: clinician communication of expectations for acupuncture effectiveness. Patient Educ Couns 2012;89(2):245–251.
31. Suarez-Almazor ME, Looney C, Liu Y, et al. A randomized controlled trial of acupuncture for osteoarthritis of the knee: effects of patient-provider communication. Arthritis Care Res (Hoboken) 2010;62(9): 1229–1236.
32. de Roode A, Heymans MW, van Lankveld W, et al. The impact of contextual effects in exercise therapy for low back pain: a systematic review and meta-analysis. BMC Med 2024;22(1):484.
33. Neogi T, Colloca L. Placebo effects in osteoarthritis: implications for treatment and drug development. Nat Rev Rheumatol 2023;19(10): 613–626.
34. Zhang W, Robertson J, Jones AC, et al. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis 2008;67(12):1716–1723.
35. Previtali D, Merli G, Di Laura Frattura G, et al. The long-lasting effects of "placebo injections" in knee osteoarthritis: a meta-analysis. Cartilage 2021;13(suppl 1):185S–196S.
36. Cashin AG, McAuley JH, Lamb SE, et al. Disentangling contextual effects from musculoskeletal treatments. Osteoarthritis Cartilage 2021;29(3):297–299.
37. McDevitt AW, O'Halloran B, Cook CE. Cracking the code: unveiling the specific and shared mechanisms behind musculoskeletal interventions. Arch Physiother 2023;13(1):14.
38. Dieppe P, Goldingay S, Greville-Harris M. The power and value of placebo and nocebo in painful osteoarthritis. Osteoarthritis Cartilage 2016;24(11):1850–1857.
39. Zou K, Wong J, Abdullah N, et al. Examination of overall treatment effect and the proportion attributable to contextual effect in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis 2016;75(11):1964–1970.
40. Zhang W, Doherty M. Efficacy paradox and proportional contextual effect (PCE). Clin Immunol 2018;186:82–86.
41. Zhang W. The powerful placebo effect in osteoarthritis. Clin Exp Rheumatol 2019;37(5 suppl 120):118–123.
42. Saueressig T, Owen PJ, Pedder H, et al. The importance of context (placebo effects) in conservative interventions for musculoskeletal pain: a systematic review and meta-analysis of randomized controlled trials. Eur J Pain 2024;28(5):675–704.
43. Poulter D, Palese A, Rodeghiero L, et al. Contextual effects in musculoskeletal pain: are we overlooking essential factors? Front Psychol 2025;16:1537242.
44. Saueressig T, Pedder H, Owen PJ, et al. Contextual effects: how to, and how not to, quantify them. BMC Med Res Methodol 2024; 24(1):35.
45. Easter C, Thompson JA, Eldridge S, et al. Cluster randomized trials of individual-level interventions were at high risk of bias. J Clin Epidemiol 2021;138:49–59.
46. Hemming K, Taljaard M. Key considerations for designing, conducting and analysing a cluster randomized trial. Int J Epidemiol 2023; 52(5):1648–1658.
47. Yang R, Carter BL, Gums TH, et al. Selection bias and subject refusal in a cluster-randomized controlled trial. BMC Med Res Methodol 2017;17(1):94.
48. Howick J, Hoffmann T. How placebo characteristics can influence estimates of intervention effects in trials. CMAJ 2018;190(30):E908–E911.
49. Evans K, Colloca L, Pecina M, et al. What can be done to control the placebo response in clinical trials? A narrative review. Contemp Clin Trials 2021;107:106503.
50. Minns Lowe CJ, Wilson MS, Sackley CM, et al. Blind outcome assessment: the development and use of procedures to maintain and describe blinding in a pragmatic physiotherapy rehabilitation trial. Clin Rehabil 2011;25(3):264–274.
51. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348(mar07 3):g1687.
52. Enck P, Klosterhalfen S, Zipfel S. Novel study designs to investigate the placebo response. BMC Med Res Methodol 2011;11(1):90.
53. Benedetti F, Maggi G, Lopiano L, et al. Open versus hidden medical treatments: the patient's knowledge about a therapy affects the therapy outcome. Prev Treat 2003;6(1). https://doi.org/10.1037/1522-3736.6.1.61a
54. Moore RA, Wiffen PJ, Eccleston C, et al. Systematic review of enriched enrolment, randomised withdrawal trial designs in chronic pain: a new framework for design and reporting. Pain 2015;156(8): 1382–1395.
55. Dieppe P. Trial designs and exploration of the placebo response. Complement Ther Med 2013;21(2):105–108.
56. Longo UG, Corradini A, Marchetti A, et al. I am afraid I will not be able to walk, that is what worries me-the experience of patients with knee osteoarthritis before total knee arthroplasty: a qualitative study. J Clin Med 2024;13(10):2878.
57. Devilly GJ, Borkovec TD. Psychometric properties of the Credibility/Expectancy Questionnaire. J Behav Ther Exp Psychiatry 2000;31(2):73–86.
58. Dworkin RH, Turk DC, Peirce-Sandner S, et al. Research design considerations for confirmatory chronic pain clinical trials: IMMPACT recommendations. Pain 2010;149(2):177–193.
59. Cook CE, Bailliard A, Bent JA, et al. An international consensus definition for contextual factors: findings from a nominal group technique. Front Psychol 2023;14:1178560.
60. Hohenschurz-Schmidt D, Vase L, Scott W, et al. Recommendations for the development, implementation, and reporting of control interventions in efficacy and mechanistic trials of physical, psychological, and self-management therapies: the CoPPS Statement. BMJ 2023; 381:e072108.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 38–46
DOI 10.1002/acr.25656

Thinking Outside the Joints: The Impact of Nonarticular Pain on Patient-Reported Outcomes in a Prospective Longitudinal Real-World Early Rheumatoid Arthritis Cohort

Charis F. Meng,[1] Marie-France Valois,[2] Yvonne C. Lee,[3] Julia D. Caci,[4] Gilles Boire,[5] Glen Hazlewood,[6] Hugues Allard-Chamard,[5] Carol Hitchon,[7] Bindee Kuriya,[8] Carter Thorne,[9] Louis Bessette,[10] Janet Pope,[11] Susan J. Bartlett,[2] and Vivian P. Bykerk,[1] on behalf of the Canadian Early Arthritis Cohort Investigators

Objective. We aimed to understand the association of nonarticular pain (NAP), regional and widespread, with patient-reported outcomes in early rheumatoid arthritis (RA).

Methods. This real-world, multicenter study observed participants with newly diagnosed active early RA (symptoms <1 year; Clinical Disease Activity Index [CDAI] score >2.8) over the first year. Participants were examined by rheumatologists and synchronously completed body pain diagrams and Patient-Reported Outcomes Measurement Information System 29 (PROMIS-29) measures at regular intervals. Participants were classified as follows: (1) no NAP, (2) regional NAP, or (3) widespread NAP. Associations between repeated measures of NAP and PROMIS-29 over the first year were estimated in separate mixed models, adjusting for sociodemographic and RA characteristics and time-varying CDAI.

Results. These 472 participants with early RA were mostly women (66%), had a mean age of 57 (SD 14) years, and had active disease (mean CDAI 27.0 [SD 14.1]). Over one-half (n = 246, 52%) reported NAP at baseline; of these, 72% (176 of 246) had regional NAP, and 28% (70 of 246) had widespread NAP. Adjusted mean change in PROMIS-29 domains were significantly worse in patients with regional versus no NAP: physical function -1.4 (95% confidence interval [CI] -2.1 to -0.7), pain interference 2.7 (95% CI 1.9–3.5), sleep disturbance 1.2 (95% CI 0.4–2.0), fatigue 2.1 (95% CI 1.2–3.1), anxiety 1.5 (95% CI 0.7–2.4), depression 1.4 (95% CI 0.5–2.2), and social participation -2.4 (95% CI -3.3 to -1.5). Associations between widespread and no NAP were larger for pain interference (5.0 [95% CI 3.7–6.4]), fatigue (3.2 [95% CI 1.7–4.8]), and social participation (-5.6 [95% CI -7.2 to -4.0]). Mean physical function, pain interference, and social participation scores were well outside the normal range.

Conclusion. NAP interferes with key aspects of well-being in early RA. Individuals with NAP experience greater pain interference and impairments in physical and social function, supporting a need for earlier identification of and interventions targeting NAP.

INTRODUCTION

Pain is an important symptom commonly experienced by patients with rheumatoid arthritis (RA). Patients with RA can have pain inside the joints and outside the joints.[1–5] Pain outside the joint or nonarticular pain (NAP) is common and persists in nearly one-third of patients with early RA, despite treatment.[4] Physician diagnoses of musculoskeletal pain, such as tendonitis, bursitis,

Clinical trial registration: Pro00016064.

Supported by the Inflammatory Arthritis Center and the Division of Rheumatology at the Hospital for Special Surgery and Division of Rheumatology. The Canadian Early Arthritis Cohort was designed and implemented by the investigators and financially supported through unrestricted research grants from Canadian affiliates of Pfizer (2007–), AbbVie (2011–); Celltrion (2023–), Sandoz (2019–), Organon (2021–), and Accord Health (2025). Previous unrestricted support was by Nordic Pharma (2022–2024), Viatris (2023), JAMP Pharma (BioJAMP) (2023), Hoffman La Roche (2011–2023), Fresenius Kabi (2021–2023), Amgen (2007–2022), Janssen (2011–2016), UCB and Bristol-Myers Squibb (2011–2018), and Medexus (2013–2022). Mentoring efforts were supported by Dr Lee's NIH grant (K24-AR-080840).

[1]Hospital for Special Surgery, Weill Cornell Medical College, New York, New York; [2]McGill University, Montreal, Quebec, Canada; [3]Northwestern University Feinberg School of Medicine, Chicago, Illinois; [4]Hospital for Special Surgery, New York, New York; [5]Université de Sherbrooke, Sherbrooke, Quebec, Canada; [6]University of Calgary, Calgary, Alberta, Canada; [7]University of Manitoba, Winnipeg, Manitoba, Canada; [8]University of Toronto, Toronto, Ontario, Canada; [9]Southlake Regional Health Centre, Centre of Arthritis Excellence, The Arthritis Program Research Group, Newmarket, Ontario, Canada; [10]Université Laval, Quebec City, Quebec, Canada; [11]University of Western Ontario, London, Ontario, Canada.

Drs Bartlett and Bykerk are co-senior authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25656).

SIGNIFICANCE & INNOVATIONS

- Nonarticular pain (NAP) in RA increases the burden of living throughout the first year of RA.
- NAP was associated with greater impairments in physical, social, and emotional health in a dose–response pattern.
- Early diagnosis and treatment of NAP may improve well-being in patients with RA.

and spinal conditions, have been associated with patient-reported NAP.[5] NAP, both regional and widespread, is associated with higher RA disease activity[5] and negatively impacts the ability to reach RA treatment goals.[4] Despite this, NAP is frequently unaddressed in RA care, and patients with RA and NAP are more likely to seek additional treatments compared to those with RA without NAP.[5]

Patients with NAP report higher mean perceived stress scores at baseline compared to those without NAP.[4] However, little is known about how NAP in RA impacts physical functioning, social functioning, vitality, mental health, and other important areas of health-related quality of life (HR-QoL).[6,7] Further, there is sparse literature looking at how NAP affects these patient-reported outcomes (PROs) in RA over time.[8] The Patient-Reported Outcome Measurement Information System 29 (PROMIS-29) reliably and precisely measures core domains of HR-QoL (physical function, pain interference, sleep problems, fatigue, anxiety and depression, and social participation) and has been validated in people with RA.[9–11] A deeper and more holistic understanding of how NAP intersects with these aspects of well-being could lead to more integrated care approaches.[6,7]

Building on our earlier work evaluating the relationship between NAP and achieving treatment targets,[4] we aimed, in this study, to evaluate the association of NAP, both regional and widespread, with PROs across core HR-QoL domains over the first year of RA diagnosis and treatment. We hypothesized that symptoms and functional impairments would worsen progressively across groups with no NAP, regional NAP, and widespread NAP.

PATIENTS AND METHODS

Participants and setting. We analyzed data from participants enrolled in the Canadian Early Arthritis Cohort (CATCH), a multicenter prospective cohort study of adults newly diagnosed with RA and seen across 18 sites (rural, suburban, and urban) in Canada (described in detail by Bykerk et al[12]). In brief, participants must be at least 18 years of age, have had joint symptoms for ≥6 weeks and ≤12 months, have more than or equal to two swollen joints or one swollen joint at the metacarpophalangeal or proximal interphalangeal joint, and one of the following: morning stiffness lasting >45 minutes, response to nonsteroidal anti-inflammatory drugs, painful joints assessed by the metatarsophalangeal squeeze test, rheumatoid factor level >20 IU, or anti–citrullinated protein antibody positivity. They are excluded or withdrawn from the cohort if an alternate diagnosis, such as psoriatic arthritis or connective tissue disease, is made. Standardized clinical assessments and PROs are collected at in-person visits with rheumatologists every three months for the first year during routine clinical visits. Treatment is at the discretion of the treating rheumatologist and is guided by treat-to-target recommendations[13]; adherence to system-level performance benchmarks has been documented.[14]

The analytic sample included participants enrolled between January 2017 and October 2023 who had completed an updated, expanded suite of PRO measures at each visit, which included a body pain diagram (BPD)[15] (Supplementary Figure 1) and PROMIS-29. We included data from participants who had completed BPDs and PROMIS-29 measures at baseline and at least one follow-up timepoint at 6 and/or 12 months (Supplementary Figure 2). Eligibility criteria included active RA disease, as defined by a Clinical Disease Activity Index (CDAI) score ≥2.8 at baseline and treatment with disease-modifying antirheumatic drugs (DMARDs), including conventional synthetic, biologic, or targeted-synthetic DMARDs within three months after study enrollment.

Written informed consent was obtained from all study participants. Research ethics boards at all study sites approved the study; this was renewed at every site on an annual basis. The trial registration number is Pro00016064.

Measures. Sociodemographic variables and RA symptom duration were provided by patients at baseline. Rheumatologists collected information on current health status, treatment, seropositivity, and C-reactive protein levels at baseline. Clinical measures obtained at each visit included the physician global assessment (MDGA) and tender and swollen joint counts. At baseline, patients completed a visual analog scale (VAS) to assess pain intensity. Time-varying PROs included the patient global assessment (PtGA), a BPD, and the PROMIS-29. Participants came for their clinic visits (baseline and 3-, 6-, and 12-month visits) with their rheumatologists and completed their study assessments synchronously at the time of their visits. They completed a BPD[15] (Supplementary Figure 1) in which they were asked to indicate nonjoint pain on the diagram in response to the following question: "Do you have pain in areas other than your

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25656.

Address correspondence via email to Charis F. Meng, MD, at mengc@hss.edu.

Submitted for publication April 27, 2025; accepted in revised form September 9, 2025.

joints?" Per our previous study,[4] patients were classified by NAP based on prespecified patterns of NAP (Supplementary Figure 1) reported in four quadrants (right upper, right lower, left upper, and left lower) and axial sections as follows: (1) none, (2) regional NAP (defined as one to three painful sections; three sections if limited to half of body [upper half, lower half, right half, or left half]), or (3) widespread NAP (defined as three to four painful sections [three sections if bilateral plus above and below the waist]). Hands and feet were excluded from these definitions because they were considered as pain in joints.

PROMIS-29 version 1.0 was collected at each study visit. It consisted of seven domains (physical function, pain interference, fatigue, sleep disturbance, anxiety, depression, and social participation [version 1.0 measures this by evaluating satisfaction with participation in social roles]) and includes an 11-point numerical rating scale for pain intensity.[16] The PROMIS is scored on a T score metric (mean 50, SD 10). The US population–based normal range is 45 to 55. Worse PROMIS symptoms (pain interference, sleep disturbance, fatigue, anxiety, and depression) are reflected by T scores ≥55; scores between 55 and 60 are described as "mild," and scores from 60 to 70 are considered as "moderate."[17] In contrast, impairments in physical function and social function are characterized[16] by scores <45. A change of two to three points is generally considered to be clinically meaningful in most domains.[18,19]

Statistical analysis. *Descriptive univariate analysis.* We compared sociodemographic and RA clinical measures, RA treatments, pain VAS scores, PtGA scores, MDGA scores, differences between PtGA and MDGA scores, and PROMIS-29 measures across pain groups at baseline. Descriptive statistics (mean and SD, median and interquartile range, and frequencies) for the total sample and groups by NAP status at baseline (no NAP, regional NAP, and widespread NAP) were calculated.

Adjusted multivariable analysis. Adjusted mean changes (and 95% confidence intervals [CIs]) in PROMIS domains by group were calculated for the groups with NAP (both regional and widespread groups) and compared to the no NAP group. Associations between repeated NAP status and PROMIS-29 domain scores over time were estimated using separate mixed-effects linear regression models comparing the three groups. Mixed-effects models have the advantage of using all available data and can account for repeated measures. Models were adjusted for baseline age, sex, education, smoking, comorbidities (Rheumatic Comorbidity Disease Index), seropositivity, and time-varying CDAI. Adjusted mean PROMIS-29 scores over the first year were also calculated.

Sensitivity analyses. Multiple sensitivity analyses were performed to assess for potential confounders, such as the presence (at baseline) of osteoarthritis, back pain or spinal conditions, and RA treatment. We performed two sensitivity analyses adjusting for RA treatment[6,7]: (1) adjusting for the use of methotrexate, advanced therapy, and oral steroids and (2) adjusting only for oral steroid use. To assess the potential association of disease activity (beyond CDAI) with our results, we also adjusted our model for C-reactive protein (CRP) level (at the same visit).

Missing data. We excluded 315 participants because of missing data. The excluded participants, compared with those in our sample, were (at baseline) younger (mean age 53 [SD 16] vs 57 [SD 14] years), more frequently female (71% vs 66%), less frequently White (58% vs 81%), less likely to have completed post-secondary education (44% vs 61%), and less likely to report back pain or spinal conditions (15% vs 21%). Fewer were treated with oral steroids in the first three months (27% vs 36%) or treated with advanced therapy in the first nine months (7% vs 13%). Baseline RA characteristics and PROMIS-29 scores were similar between included and excluded cohort participants. Data analyses were conducted using SAS software, version 9.4, for Windows.[20]

RESULTS

A total of 472 adults with early RA were included (Supplementary Figure 2). Participants were mainly female (66%) and White (81%), with a mean age of 57 (SD 14) years and a mean symptoms duration of 5.2 (SD 2.8) months. Seventy-two percent were seropositive. They had active RA at baseline, with a mean CDAI of 27.0 (SD 14.1), and most (79%) were treated with methotrexate at a mean weekly dose of 20.6 (SD 4.0) mg.

Patterns associated with baseline NAP. Over one-half of patients reported NAP at baseline (n = 226, 52%); of these with NAP, 72% (176 of 246) had regional NAP, and 28% (70 of 246) had widespread NAP. Figure 1 shows prevalence in the entire analytic sample.

The widespread NAP group had a significantly higher Rheumatic Disease Comorbidity Index score compared to that of the no NAP group ($P < 0.01$) (see Table 1). Back pain or spinal conditions were more frequently reported in those with both regional and widespread NAP versus those with no NAP ($P = 0.001$). Fibromyalgia was reported more frequently in those with widespread NAP versus those with no NAP ($P = 0.001$) but did not significantly differ between regional and no NAP groups. Participants with widespread NAP had higher mean CDAI scores and CRP levels compared to the other groups, although differences did not reach statistical significance. A significant dose–response relationship with PROs was observed, with higher pain VAS scores ($P < 0.0001$) and PtGA scores ($P < 0.0001$), corresponding with increasing NAP distribution. The same dose–response pattern was observed with PROMIS domains, in which symptoms and function worsened as NAP distribution increased ($P < 0.0001$). Patients without NAP reported scores within normal limits (ie, 45–55) for fatigue, anxiety, depression, sleep disturbance, and social participation, with mild impairments in physical

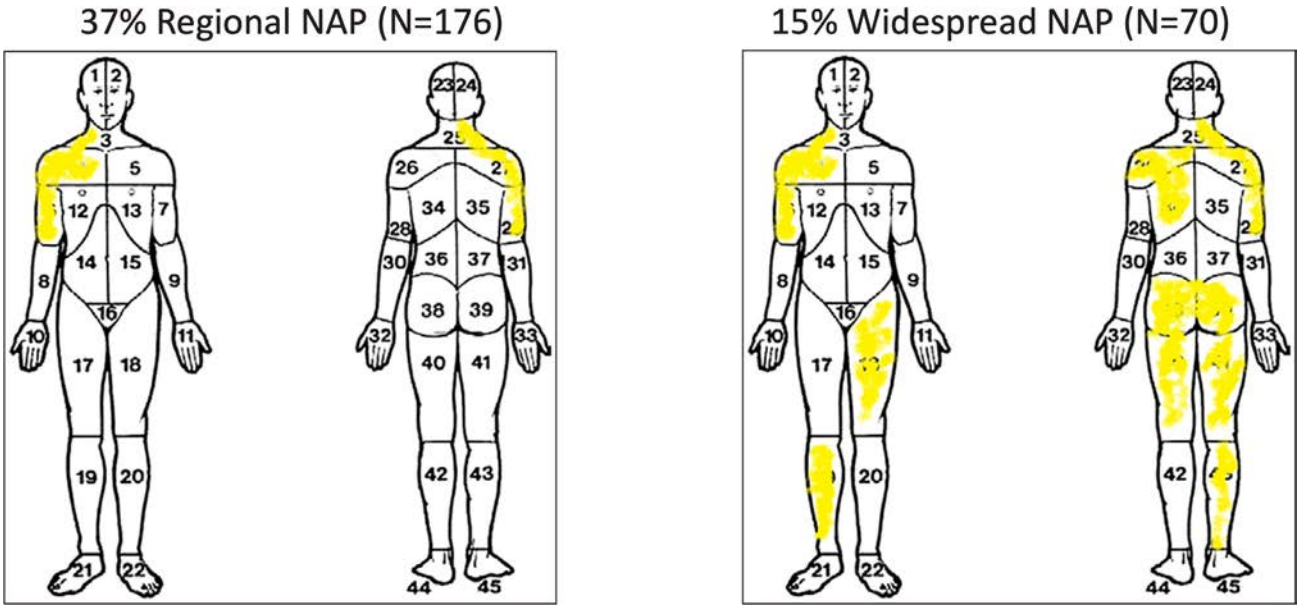


Figure 1. Baseline prevalence of regional and widespread NAP in early RA cohort (N = 472). Figures are illustrative examples of regional and widespread NAP, adapted from the body pain diagram by Margolis et al.[15] Prevalence is calculated for the total analytic sample (N = 472). NAP, nonarticular pain; RA, rheumatoid arthritis.

and social function and mild pain interference (see Table 1). Patients with regional NAP reported moderate impairments in physical and social function and moderate pain interference, along with mild fatigue and sleep disturbance. Widespread NAP was associated with elevated scores in all domains, including moderate impairments in physical and social function, moderate pain interference, and mild fatigue and sleep disturbance. Notably, anxiety and depression were within normal limits for all except those with widespread NAP, who reported mild symptom levels at baseline.

Associations of NAP with symptoms and function over the first year. In multivariable mixed-effects regression models, compared to the no NAP group, both the regional and widespread NAP groups reported significantly worse physical function, pain interference, sleep disturbance, fatigue, anxiety, depression, and social participation, with greater effects seen with widespread pain (see Table 2). The largest mean changes over the first year were reported by the widespread NAP group for pain interference (5.0 [95% CI 3.7–6.4]), fatigue (3.2 [95% CI 1.7–4.8]), and social participation (-5.6 [95% CI -7.2 to -4.0]). These changes have been identified by patients with RA as being clinically meaningful.[18,19] Mean changes over the first year for the regional NAP group ranged from 1 to 2 points, except for social participation (-2.4) and pain interference (2.7).

Figure 2 presents adjusted mean T scores for PROMIS-29 domains over the first year. Across all groups, greater NAP distribution was associated with worse symptoms and function. All groups reported impairments in physical and social function and pain interference outside the normal range. The no NAP group had scores within the normal range for all other domains, whereas the regional NAP group reported mild fatigue, with sleep disturbance, anxiety, and depression within normal limits. The group with widespread NAP reported mild fatigue, sleep disturbance, and anxiety. Mean depression scores for all groups were within the normal range.

Sensitivity analyses. The results of our analyses adjusting for RA treatment in both versions (adjusting for methotrexate, advanced therapies, and oral steroids and adjusting for oral steroids only) were similar to those seen in our primary analysis (Supplementary Table 1). Adjusting for CRP level slightly increased the associations between NAP and PROMIS outcomes. Results were also robust to adjusting for presence of baseline osteoarthritis and back pain and spinal conditions (Supplementary Table 2). All results remained statistically significant.

DISCUSSION

This is the first prospective, longitudinal, multicenter, real-world early RA cohort study to show that having NAP adversely affects physical functioning, social functioning, fatigue, and other essential aspects of well-being throughout the first year of RA. In support of our hypothesis, we saw a dose–response relationship between NAP and impairment in HR-QoL across seven domains of health. Widespread NAP was associated with significantly greater and clinically meaningful impacts on symptoms and

Table 1. Baseline Characteristics across Non-Articular Pain Patterns in Early RA Cohort.

	No NAP (N = 226)	Regional NAP (N = 176)	Widespread NAP (N=70)	*P* value[β]
Age (year), mean (SD)	58 (14)	56 (14)	56 (14)	0.42
Female, n (%)	146 (65)	118 (67)	49 (70)	0.68
White, n (%)	174 (77)	149 (85)	57 (81)	0.34
Body Mass Index (kg/m^2) mean (SD)	28.0 (6.9)	28.7 (6.9)	29.6 (7.5)	0.26
Obese, n (%)[Φ]	70 (33)	52 (32)	26 (39)	0.50
RDCI (0-9), mean (SD)	**1.2 (1.3)$_a$**	**1.4 (1.3)$_{a,b}$**	**1.8 (1.6)$_b$**	**<0.01**
Current smoker, n(%)	29 (13)	27 (15)	14 (20)	0.33
Postsecondary education, n (%)	130 (58)	118 (67)	42 (60)	0.27
Fibromyalgia*, n (%)	**2 (1)$_a$**	**4 (2)$_a$**	**7 (10)$_b$**	**0.001**
Osteoarthritis*, n (%)	68 (30)	46 (26)	20 (29)	0.68
Back/spine condition*, n (%)	**33 (15)$_a$**	**45 (26)$_b$**	**23 (33)$_b$**	**0.001**
RA symptom duration (months), mean (SD)	5.2 (2.8)	5.3 (2.9)	5.1 (2.5)	0.94
Seropositivity (RF/anti-CCP), n (%)[Φ]	**169 (78)$_a$**	**109 (65)$_b$**	**46 (71)$_{a,b}$**	**0.02**
CDAI, mean (SD)	26.6 (14.9)	26.4 (12.8)	29.4 (14.5)	0.29
Tender joint count (TJC)28, median (IQR)	7 (3, 12)	8 (4, 12)	9 (4, 13)	0.48
Swollen joint count (SJC)28, median (IQR)	7 (3, 11)	7 (3, 11)	7 (3.5, 10.5)	0.62
TJSD (TJC-SJC), mean (SD)	0.7 (4.8)	1.1 (4.6)	1.7 (5.0)	0.30
C-reactive protein (mg/L), median (IQR)	7.0 (3.0, 19.5)	5.8 (2.5, 15.5)	8.5 (1.9, 22.4)	0.16
Treatment[Δ]				
Oral Steroids, n (%)	82 (36)	55 (31)	21 (30)	0.46
Methotrexate, n (%)	177 (78)	137 (78)	58 (83)	0.66
Methotrexate dose (mg/week), mean (SD)	20.1 (4.2)	21.2 (3.7)	20.9 (3.9)	0.05
Advanced therapies[¥] (first 9 months), n (%)	26 (12)	21 (12)	15 (21)	0.08
Pain Visual Analog Scale (0-10), mean (SD)	**4.8 (3.0)$_a$**	**6.1 (2.5)$_b$**	**7.0 (2.4)$_c$**	**<0.0001**
Patient global assessment (PtGA), mean (SD)	**4.1 (2.8)$_a$**	**5.2 (2.6)$_b$**	**6.1 (2.5)$_c$**	**<0.0001**
Physician global assessment (MDGA), mean (SD)	5.6 (2.6)	5.3 (2.3)	5.6 (2.2)	0.44
Difference PTGA-MDGA, mean (SD)	**−1.5 (3.4)$_a$**	**-0.1 (3.2)$_b$**	**0.5 (3.0)$_b$**	**<0.0001**
PROMIS29 T-scores, mean (SD)				
Physical function	**42.4 (8.3)$_a$**	**39.9 (6.9)$_b$**	**38.1 (8.2)$_b$**	**<0.0001**
Pain interference	**58.2 (8.8)$_a$**	**61.8 (7.0)$_b$**	**64.4 (8.7)$_b$**	**<0.0001**
Sleep disturbance	**51.7 (8.5)$_a$**	**55.0 (8.7)$_b$**	**56.7 (7.4)$_b$**	**<0.0001**
Fatigue	**51.4 (10.4)$_a$**	**55.6 (9.6)$_b$**	**58.4 (10.7)$_b$**	**<0.0001**
Anxiety	**50.8 (9.7)$_a$**	**53.8 (9.6)$_b$**	**56.2 (11.0)$_b$**	**<0.0001**
Depression	**49.6 (9.5)$_a$**	**52.6 (9.2)$_b$**	**56.0 (11.1)$_c$**	**<0.0001**
Social participation	**45.9 (9.9)$_a$**	**42.2 (9.0)$_b$**	**37.8 (9.1)$_c$**	**<0.0001**

% of non-missing; *Reported by MD or patient; [Δ] Treatment at baseline unless otherwise indicated. [¥]Refers to TNF-inhibitors, biosimilars (both TNF and non-TNF inhibitors), non-TNF-inhibitor biologic DMARDs and Janus kinase inhibitors. RDCI Rheumatic Disease Comorbidity Index; CDAI Clinical Disease Activity Index; csDMARD conventional synthetic disease modifying anti-rheumatic drug.
[β] *P* value for overall difference from an ANOVA model for continuous variables and from a chi-square test for categorical variables or Fisher's exact test in case of categorical variables with cells smaller than 5. Differing subscripts indicate significantly different values ($P < 0.05$).

function, with the greatest associations observed with pain interference, fatigue, and the ability to participate in social roles. Significant but smaller associations were seen with regional NAP. Further, both regional and widespread NAP were associated with impairments in multiple aspects of HR-QoL, especially with respect to physical function, pain interference, and social participation.

Evidence suggesting a dose–response of NAP (none, regional, and widespread) was observed in all seven domains of PROMIS-29. Compared to those without NAP, the regional pain group reported somewhat worse symptoms and impairment, and the widespread pain group reported meaningfully worse symptoms and function. The most substantial impairments were noted in pain interference, fatigue, and social participation. This is important because a systematic review of early inflammatory arthritis concluded that baseline PROs, including pain, fatigue, and physical function (and not clinical indicators or imaging), were the most robust predictors of disability at five years.[21] Additionally, patients with RA have rated fatigue as a high-priority symptom that often goes unaddressed, which impacts their ability to participate in social roles and activities.[22] Hence, it is not surprising that, on average, participants with NAP reported worse pain interference and fatigue and impairments in physical and social function. There is growing awareness that social isolation and loneliness are independently associated with adverse health effects, particularly in aging adults,[23] which is relevant because our cohort primarily consisted of middle-aged women. Social participation, a modifiable determinant of health, may help to mitigate these risks.[23] Compared with the general population, early RA is already associated with significant impairments in physical function, social participation, and pain interference.[18] Importantly, our data suggest NAP is associated with a greater burden of living

Table 2. Multivariate mixed-effects regression analysis showing associations of NAP with mean changes in PROMIS-29 T scores and mean T scores in early RA cohort*

PROMIS-29 outcomes	Mean change[a]	95% CI	T score[b]	95% CI
Physical function (baseline average of 40.8)				
No NAP	Reference group	–	40.8	–
Regional NAP	−1.4	−2.1 to −0.7	39.4	38.7 to 40.1
Widespread NAP	−3.0	−4.2 to −1.8	37.8	36.6 to 39.0
Pain interference (baseline average of 60.4)				
No NAP	Reference group	–	60.4	–
Regional NAP	2.7	1.9 to 3.5	63.1	62.3 to 63.9
Widespread NAP	5.0	3.7 to 6.4	65.4	64.1 to 66.8
Sleep disturbance (baseline average of 53.7)				
No NAP	Reference group	–	53.7	–
Regional NAP	1.2	0.4 to 2.0	54.9	54.1 to 55.7
Widespread NAP	2.7	1.3 to 4.1	56.4	55.0 to 57.8
Fatigue (baseline average of 54.0)				
No NAP	Reference group	–	54.0	–
Regional NAP	2.1	1.2 to 3.1	56.1	55.2 to 57.1
Widespread NAP	3.2	1.7 to 4.8	57.2	55.7 to 58.8
Anxiety (baseline average of 52.7)				
No NAP	Reference group	–	52.7	–
Regional NAP	1.5	0.7 to 2.4	54.2	53.4 to 55.1
Widespread NAP	2.7	1.2 to 4.2	55.4	53.9 to 56.9
Depression (baseline average of 51.7)				
No NAP	Reference group	–	51.7	–
Regional NAP	1.4	0.5 to 2.2	53.1	52.2 to 53.9
Widespread NAP	3.0	1.5 to 4.4	54.7	53.2 to 56.1
Social participation (baseline average of 43.3)				
No NAP	Reference group	–	43.3	–
Regional NAP	−2.4	−3.3 to −1.5	40.9	40.0 to 41.8
Widespread NAP	−5.6	−7.2 to −4.0	37.7	36.1 to 39.3

* Fully adjusted multivariate mixed-effects models adjusted for time and baseline variables of age, sex, education, current smoker, Rheumatic Disease Comorbidity Index, seropositive status, and time-variant Clinical Disease Activity Index (at same visit as outcome). CI, confidence interval; NAP, nonarticular pain; PROMIS-29, Patient-Reported Outcomes Measurement Information System 29; RA, rheumatoid arthritis.
[a] Change in T scores averaged over time.
[b] T score from baseline average T score + NAP change over first year of follow-up.

with RA. These findings support the importance of early identification of NAP and targeted interventions to treat NAP. This may help patients remain more fully able to engage in activities and roles that improve overall quality of life.

Chronic pain and fatigue are often viewed as hallmarks of fibromyalgia. However, fibromyalgia is also associated with a higher likelihood of mood disorders and other symptoms.[24] We found that anxiety and depression scores were only slightly higher in those with NAP and generally within normal limits. A potential explanation for this could be that our definition of widespread NAP had a minimum of three sections in an extensive distribution on the BPD. The rationale behind our decision was based on data showing that patients with three sections share clinical features more closely resembling those of patients with four sections than those of patients with one or two sections.[4] Thus, using our widespread NAP group, we included participants with less extensive pain distribution compared to those meeting more extensive distribution criteria used by other investigators. Taken together, these findings lend support to the notion that although NAP may share some overlap with symptoms of fibromyalgia, an important characteristic of NAP as defined in this cohort may be the elevation of pain, fatigue, and functional limitations in the absence of an affective component. These observations may be considered in the clinic and could help increase clinician awareness of NAP in patients with RA.

Our primary analysis did not include adjusting for RA treatment for the sake of parsimony and because RA therapies have not yet been proven to affect the classification of NAP. Further, our sensitivity analysis showed that adjusting for treatment did not change our results. We have previously shown that NAP is associated with disease activity as assessed by CDAI.[5] Our goal for the current study was to evaluate the impact of NAP when dissociated from disease activity. Even after adjustment for CDAI, our results showed an impact of NAP on PROs. There could be several explanations for why the magnitude of the association was reduced when adjusted for CDAI. Disease activity could contribute to the impact of NAP on PROs. Another explanation is that NAP may be driving components of CDAI (such as the PtGA) higher. If this is the case, adjusting for CDAI may underestimate the true magnitude of associations of NAP and PROs. A study

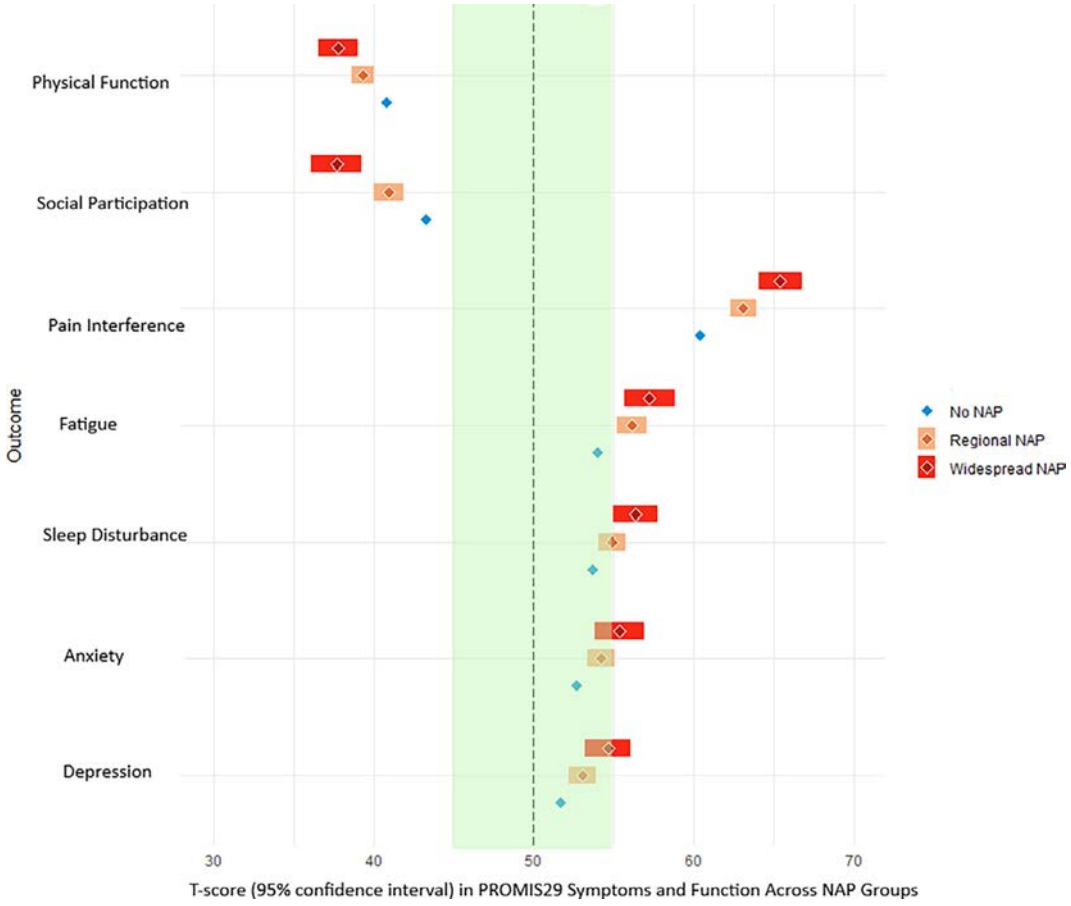


Figure 2. The association of NAP and mean T scores for PROMIS29 symptoms and function from adjusted mixed-effects regression analysis. T score is calculated from the baseline average T score + the adjusted mixed-effects regression coefficient for NAP, with all other covariates in the model remaining the same. Normal range of T scores (45–55) is indicated by the light green shaded area. NAP, nonarticular pain; PROMIS29, Patient-Reported Outcomes Measurement Information System 29.

looking at these questions is currently underway. Interestingly, adjusting for CRP level increased the association of NAP with PROMIS measures. It is possible that in the absence of measured systemic inflammation, NAP has a greater effect on PROs.

Our results are consistent with prior studies. A cross-sectional study of 38 patients with RA reported that those who reported a higher number of painful body sites were found to have greater activity limitations, more severe sleep disturbances, and higher pain intensity.[25] A longitudinal study of patients with early RA showed that those with widespread nonjoint pain had worse HR-QoL, as assessed by the 36-item Short Form Health Survey at the three-year follow-up.[8] Both studies focused on widespread NAP. Our report extends their findings by including an evaluation of both regional and widespread NAP. Compared to widespread pain, regional NAP was more common, and we have shown previously that it is also more likely to persist or worsen in the first year of RA.[4]

Strengths of our study include its longitudinal design, early diagnosis, and treat-to-target paradigm of management in a large well-defined cohort of participants with early RA seen in routine care. This supports the external validity of our findings to many patients with RA. PROMIS-29 has been shown to be more reliable, precise, and responsive than previous HR-QoL measures.[9,11] Additional strengths of our study include the consideration of the effect of multiple confounders tested in several sensitivity analyses and adjusting for potential confounders in our models. These results remained similar to those seen in our primary analysis. Mixed-effects modeling allowed us to use available data from all patients at all time points.

Our study has limitations. This study is observational, and our results are subject to residual unmeasured confounding. Most of our participants were White. Our study's large sample size and multicenter design included urban, suburban, and rural sites across Canada, where patients have similar access to care. We did not adjust for a fibromyalgia diagnosis given the high correlation between fibromyalgia and NAP[24] and the low (3%) baseline prevalence of fibromyalgia. The presence of fibromyalgia, osteoarthritis, and back pain or spinal conditions was reported either by the patients, doctors, or both and may have been underreported. There is potential for multiple comparisons bias. We

chose not to correct for multiple comparisons given that this is an exploratory study. Our study overlapped with the COVID-19 pandemic, which limited some in-person assessments for CATCH at a time when the BPD was added to our suite of assessments. However, all our study sites experienced the same constraints. Although this may have decreased the precision of our results, it is unlikely to have introduced bias. Excluded participants were less likely to report back pain or spinal conditions, which may have affected our study estimates. However, our sensitivity analyses adjusting for presence of back pain or spinal conditions showed similar results to those of our primary analysis. Excluded participants and included participants had similar baseline PROMIS scores. PROMIS measures capture an assessment of pain over the past seven days and therefore may not reflect important fluctuations that occur between assessment points.

In conclusion, pain outside the joints is associated with greater pain and fatigue and greater impairments in physical and social function. NAP may also occur with mild or absent mood symptoms. Our findings can help increase clinician awareness of NAP in patients with RA. Early recognition may enable opportunities to provide appropriate treatment in addition to DMARD therapy (eg, mobilization, glucocorticoid injections, sleep hygiene, and pain management). More work is needed to validate the use of BPDs in patients with RA, and further investigation of the potential mechanisms underlying NAP is warranted. This should be followed by interventional studies that can be readily implemented in clinical care to target NAP before it becomes chronic and affects overall disease status and patient quality of life.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Hammer HB, Kvien TK, Terslev L. Tenosynovitis in rheumatoid arthritis patients on biologic treatment: involvement and sensitivity to change compared to joint inflammation. Clin Exp Rheumatol 2017;35(6):959–965.
2. Hammer HB, Kvien TK, Terslev L. Intermetatarsal bursitis is frequent in patients with established rheumatoid arthritis and is associated with anti-cyclic citrullinated peptide and rheumatoid factor. RMD Open 2019;5(2):e001076.
3. Hagege B, Tubach F, Alfaiate T, et al. Increased rate of lumbar spondylolisthesis in rheumatoid arthritis: A case-control study. Eur J Clin Invest 2018;48(9):e12991.
4. Meng CF, Caci J, Schieir O, et al. Characterizing non-articular pain at early rheumatoid arthritis diagnosis, evolution over the first year of treatment and impact on remission in a prospective real-world early rheumatoid arthritis cohort. Arthritis Rheumatol 2025;77(4):405–413.
5. Meng C, Caci J, Jannat-Khah D, et al. The association of patient-reported non-articular pain with musculoskeletal pain diagnoses and RA disease activity in a prospective real-world cohort of patients with early RA [abstract]. Arthritis Rheumatol 2024;76(suppl 9). https://acrabstracts.org/abstract/the-association-of-patient-reported-non-articular-pain-with-musculoskeletal-pain-diagnoses-and-ra-disease-activity-in-a-prospective-real-world-cohort-of-patients-with-early-ra
6. Zhong Y, Lai R, Deng X. Beyond joint pain: expanding insights on the impact of nonarticular pain in early rheumatoid arthritis: comment on the article by Meng et al. Arthritis Rheumatol Published online March 19, 2025. https://doi.org/10.1002/art.43160
7. Meng CF, Lee YC, Bartlett SJ, et al. Reply. Arthritis Rheumatol Published online April 6, 2025. https://doi.org/10.1002/art.43158
8. Schelin M, Westerlind H, Lindqvist J, et al. Widespread non-joint pain in early rheumatoid arthritis. Scand J Rheumatol 2021;50(4):271–279.
9. Bartlett SJ, Orbai AM, Duncan T, et al. Reliability and validity of selected PROMIS measures in people with rheumatoid arthritis. PLoS One 2015;10(9):e0138543.
10. Bartlett SJ, De Leon E, Orbai AM, et al. Patient-reported outcomes in RA care improve patient communication, decision-making, satisfaction and confidence: qualitative results. Rheumatology (Oxford) 2020;59(7):1662–1670.
11. Orbai AM, Bingham CO III. Patient reported outcomes in rheumatoid arthritis clinical trials. Curr Rheumatol Rep 2015;17(4):28.
12. Bykerk VP, Jamal S, Boire G, et al. The Canadian Early Arthritis Cohort (CATCH): patients with new-onset synovitis meeting the 2010 ACR/EULAR classification criteria but not the 1987 ACR classification criteria present with less severe disease activity. J Rheumatol 2012;39(11):2071–2080.
13. Bykerk VP, Akhavan P, Hazlewood GS, et al; Canadian Rheumatology Association. Canadian Rheumatology Association recommendations for pharmacological management of rheumatoid arthritis with traditional and biologic disease-modifying antirheumatic drugs. J Rheumatol 2012;39(8):1559–1582.
14. Barber CEH, Schieir O, Lacaille D, et al; Canadian Early Arthritis Cohort Investigators. High adherence to system-level performance measures for rheumatoid arthritis in a national early arthritis cohort over eight years. Arthritis Care Res (Hoboken) 2018;70(6):842–850.
15. Margolis RB, Chibnall JT, Tait RC. Test-retest reliability of the pain drawing instrument. Pain 1988;33(1):49–51.
16. Nagaraja V, Mara C, Khanna PP, et al. Establishing clinical severity for PROMIS® measures in adult patients with rheumatic diseases. Qual Life Res 2018;27(3):755–764.
17. HealthMeasures. PROMIS score cut points. Northwestern University. https://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/promis-score-cut-points. Accessed August 11, 2025.
18. Bartlett SJ, Gutierrez AK, Andersen KM, et al. Identifying minimal and meaningful change in a patient-reported outcomes measurement information system for rheumatoid arthritis: use of multiple methods and perspectives. Arthritis Care Res (Hoboken) 2022;74(4):588–597.
19. Terwee CB, Peipert JD, Chapman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 2021;30(10):2729–2754.
20. Base SAS 9.4 Procedures Guide: Statistical Procedures. 4th ed. SAS Institute Inc; 2015.
21. Gwinnutt JM, Sharp CA, Symmons DPM, et al. Baseline patient reported outcomes are more consistent predictors of long-term functional disability than laboratory, imaging or joint count data in patients with early inflammatory arthritis: a systematic review. Semin Arthritis Rheum 2018;48(3):384–398.
22. Choy EH, Dures E. Fatigue in rheumatoid arthritis. Rheumatology (Oxford) 2019;58(suppl 5):v1–v2.

23. Dawson-Townsend K. Social participation patterns and their associations with health and well-being for older adults. SSM Popul Health 2019;8:100424.
24. Wolfe F, Clauw DJ, Fitzcharles MA, et al. 2016 Revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46(3):319–329.
25. Aydemir B, Song J, Muhammad L, et al. Association between activity limitations, sleep disturbance, pain sensitivity, pain intensity, and widespread pain in adults with rheumatoid arthritis [abstract]. Arthritis Rheumatol 2022;74(suppl 9). https://acrabstracts.org/abstract/association-between-activity-limitations-sleep-disturbance-pain-sensitivity-pain-intensity-and-widespread-pain-in-adults-with-rheumatoid-arthritis/

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 47–53
DOI 10.1002/acr.25668

American College of Rheumatology
Empowering Rheumatology Professionals

Association Between Discordance of Disease Activity Indices and Quantitative Sensory Testing Measures of Nociplastic Pain in Patients With Rheumatoid Arthritis

Clarice P. Lin,[1] Burcu Aydemir,[1] Jing Song,[1] Lutfiyya N. Muhammad,[1] Tuhina Neogi,[2] Wendy Marder,[3] Clifton O. Bingham,[4] Marcy B. Bolster,[5] Daniel J. Clauw,[3] Dorothy D. Dunlop,[1] and Yvonne C. Lee[1]

Objective. This study investigates the association between discordance in commonly collected clinical indicators of rheumatoid arthritis (RA) disease activity and abnormalities in quantitative sensory testing (QST) observed in individuals with nociplastic pain. The goal is to identify low-burden methods of assessing nociplastic pain in rheumatology practice.

Methods. Data from 225 patients with active RA were included for cross-sectional analyses. Measures of discordance in disease activity were as follows: (1) tender–swollen joint count difference (TSJD), (2) proportion of subjective components to the total Disease Activity Score in 28 joints (DAS28-P), and (3) patient global assessment of disease activity minus evaluator global assessment of Disease Activity (PtGA – EGA). QST measures were pressure pain thresholds (PPTs) at the trapezius, temporal summation (TS), and conditioned pain modulation. We evaluated associations between measures of discordance and QST using unadjusted and multivariable linear regression models.

Results. The mean TSJD was 5.4 (SD ±8.2), and the mean DAS28-P was 49.7% (SD ±13.3%). The mean PtGA – EGA was 0.7 (SD ±2.2). Higher TSJD was associated with lower trapezius PPT (β = –0.05; 95% confidence interval [CI] –0.08 to –0.02) and higher TS (β = 0.29; 95% CI 0.05 to 0.53). Higher DAS28-P was associated with a lower trapezius PPT (β = –0.05; 95% CI –0.07 to –0.04) and higher TS (β = 0.21; 95% CI 0.06 to 0.35). PtGA – EGA was not associated with any QST measures.

Conclusion. Two of our proposed measures of discordance (higher TSJD and DAS28-P) were modestly associated with worse QST measures of nociplastic pain (lower trapezius PPT and higher TS), suggesting that discordance between patient-reported and physician-assessed measures of disease activity may reflect an element of nociplastic pain.

INTRODUCTION

Pain is a predominant symptom of rheumatoid arthritis (RA), with over 71% of patients citing pain as a priority in their care.[1] Patients with RA may experience nociceptive pain due to active joint inflammation; however, they may also experience other types of pain, which can prove challenging for rheumatology providers to assess and treat.

Abnormalities in central nervous system pain processing pathways can lead to nociplastic pain, which refers to pain that is not related to tissue lesion or disease or injury.[2] Nociplastic pain is generally suspected in patients with RA who have pain, despite normalization of markers of inflammation and reduced swollen joint counts.[3] Structural damage to joints may also contribute to elevated RA indices.[4] Noninflammatory pain may lead to elevated assessments of RA disease activity, result in overtreatment, and

Supported by the Camille Julia Morgan Arthritis Research and Education Fund and the National Institutes of Health (grants R01-AR-064850, K24-AR-080840, P30-AR-072579, P30-AR-070254, K24-AR-070892, P30-AR-072571, and T32AR085530).

[1]Northwestern University Feinberg School of Medicine, Chicago, Illinois; [2]Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts; [3]University of Michigan, Ann Arbor; [4]Johns Hopkins University School of Medicine, Baltimore, Maryland; [5]Massachusetts General Hospital, Boston.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25668).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25668.

Address correspondence via email to Clarice P. Lin, MD, at clarice.lin@nm.org.

Submitted for publication April 30, 2025; accepted in revised form September 23, 2025.

SIGNIFICANCE & INNOVATIONS

- This study analyzed possible associations between discordance in commonly collected clinical indicators of rheumatoid arthritis (RA) disease activity and abnormalities in quantitative sensory testing (QST) observed in individuals with nociplastic pain.
- Modest associations between tender–swollen joint count difference (TSJD), the proportion of subjective components to the total Disease Activity Score in 28 joints (DAS28-P), and QST measures suggest that TSJD and DAS28-P could provide information on the presence of nociplastic pain in patients with RA.
- We favor TSJD over DAS28-P in clinical practice for identifying these patients, given its practicality in application.

be associated with poor clinical outcomes.[5] To develop more effective pain management strategies, a deeper comprehension is required to understand noninflammatory pain in patients with RA.

In research settings, quantitative sensory testing (QST) is a method used to assess alterations in nociceptive signaling, which may reflect nociplastic pain.[6] QST uses different experimental stimuli to elicit and quantitatively measure responses to painful and nonpainful stimuli. QST measures include pressure pain thresholds (PPTs) at the trapezius, temporal summation (TS), and conditioned pain modulation (CPM). Despite their use in research settings, QST measures are challenging to implement in the clinical environment, given the need for rigorous training, time, and controlled settings. These measures have historically been considered the gold standard for diagnosing central sensitization and then nociplastic pain, although these measures are only abnormal in a subset of individuals with conditions such as fibromyalgia.[7–9]

Previous studies have proposed several measures of discordance between patient-reported and physician-assessed measures of disease activity as potential ways to identify patients with RA and fibromyalgia, the prototypical nociplastic pain condition. Pollard et al found that patients with RA and concomitant fibromyalgia had a significantly larger tender–swollen joint count difference (TSJD) on the standardized 28-joint examination, compared to those with RA but no fibromyalgia.[10] McWilliams et al demonstrated that a high proportion of subjective components to the total Disease Activity Score in 28 joints (DAS28-P), suggests higher levels of nociplastic pain, whereas Salaffi et al suggested that the DAS28-P showed good discriminatory power in distinguishing patients with RA who have fibromyalgia from those without fibromyalgia.[11,12] Studenic et al showed that around 80% of the patient global assessment of disease activity (PtGA) is explained by pain, whereas the swollen joint count explained the majority of the evaluator global assessment of disease activity (EGA), thereby suggesting PtGA – EGA as a potential measure for identifying discrepancies between pain and inflammation.[13]

There is a critical need to measure and detect nociplastic pain in patients with RA to decrease pain burden and improve treatment, but there is currently no low-burden tool used in clinical practice. The components of TSJD, DAS28-P, and PtGA – EGA are already routinely assessed in clinical practice; however, their associations with research-based assessments of pain sensitization that may reflect nociplastic pain in patients with RA are unknown. The purpose of this study was to evaluate the relationship between QST measures of pain sensitization with clinical measures of disease discordance. We hypothesized that larger discordance in disease activity measures would be associated with greater evidence of nociplastic pain, defined by lower PPT, higher TS, and lower CPM. If these measures were associated, rheumatology health care providers might be able to incorporate these assessments into their workflow, leading to a better understanding of the impact of nociplastic pain on patients' experience of disease and, ultimately, improved health outcomes.

PATIENTS AND METHODS

Study population. The Central Pain in Rheumatoid Arthritis (CPIRA) study is a prospective, observational cohort study conducted at five academic medical centers, including Brigham and Women's Hospital, Massachusetts General Hospital, University of Michigan Medical Center, Johns Hopkins University, and Boston University School of Medicine.[14,15] Participants were recruited between 2014 and 2017. Enrollment criteria included the following: (1) aged ≥18 years old, (2) meets American College of Rheumatology (ACR)/EULAR 2010 criteria for RA,[16] and (3) started on disease-modifying antirheumatic drugs (DMARDs) or switched DMARDs because of lack of efficacy.[17] Participants taking centrally acting pain medications were required to be taking stable doses for at least three months before study entry and expected to be taking stable doses for the duration of the study. Patients with known peripheral neuropathy were excluded. Institutional review board approval was obtained at each institution, and all participants provided written informed consent. For the purposes of this specific study, we included 225 participants with complete baseline data in outcome and covariates for cross-sectional data analysis.

Clinical assessments and patient-reported outcomes. Study participants also underwent a standardized assessment of tenderness and swelling at 28 joints by trained assessors. Blood was drawn to assess C-reactive protein (CRP) level, a measure of inflammation. The DAS28 was calculated using the tender joint count, swollen joint count, PtGA, and CRP level.[11] The Clinical Disease Activity Index was calculated by adding the 28-tender joint count, 28-swollen joint count, PtGA, and EGA.[18] Smoking status was obtained via self-report

questionnaires administered during clinical visits. Current pain intensity was assessed using a 0-to-10 numerical rating scale. Patients also completed the Patient-Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests to assess pain interference, depression, and sleep disturbance.[5,19–21] Fibromyalgia was defined according to the ACR 2016 revised fibromyalgia diagnostic criteria.[22]

QST. QST measures included PPT at the trapezius, TS, and CPM. A brief description of each measure is provided in the following sections. Detailed descriptions have been published previously.[23]

PPT at the trapezius. PPTs at the trapezius (a site distant from inflamed joints) reflect overall pain sensitivity. Decreased PPTs at nonarticular sites are thought to be indicative of nociceptive disruption and pain sensitization.[24] A pressure algometer (Wagner Force 10 FDX) was used to assess PPTs at the center of the trapezius muscle. Pressure was gradually increased until patients indicated that the stimulus was first perceived as painful. Possible PPT values ranged from 0 to 11 kilogram-force (kgf). Lower values indicate greater pain sensitivity.

TS of the forearm. TS is considered a measure of ascending pain facilitation; a high TS is thought to reflect increased responsiveness of the dorsal horn neurons to peripheral stimulation, which is suggestive of nociplastic pain.[25,26] TS was evaluated by tapping a calibrated probe (MRC Systems) on the patient's forearm. Probes of increasing weight were used to identify the probe that elicited a pain level of 30 to 40 on a 0 to 100 scale. This probe was then tapped 10 times on the forearm. Patients were asked to rate the pain intensity of the first and last stimulus. TS was calculated as the difference between pain intensity at the first and last stimulus. Possible TS values ranged from 0 to 100. Higher values indicate greater pain facilitation.

CPM. CPM reflects descending pain inhibition, another proposed mechanism of nociplastic pain. Low CPM suggests decreased activity of the descending analgesic pathways, resulting in inefficient pain inhibition and increased nociplastic pain.[26,27] To determine CPM, the participant's right hand was immersed in a 5- to 7-°C water bath. The PPT at the left trapezius was measured before and after immersion. CPM was calculated as the ratio of the PPT at the trapezius after immersion to the PPT at the trapezius before immersion. Because CPM is a ratio, the range of possible values is infinite, although typical values range from 0 to 4. Low CPM suggests decreased activity of the descending analgesic pathways, inefficient pain inhibition, and, thus, increased nociplastic pain.

Discordance measures. The TSJD was calculated by subtracting the 28-swollen joint count from the 28-tender joint count.[10] A positive TSJD indicates a greater number of tender joints relative to swollen joints. A previous study reported that a TSJD ≥7 identifies patients with fibromyalgic RA with 83% sensitivity and 80% specificity.[10]

The DAS28-P, which is the ratio of subjective components of the DAS28 (tender joint count and patient global assessment) to the total DAS28, was calculated using the following equation[11]:

$$\frac{\left(\left(0.56 * \sqrt{TJC28} + (0.014 * Patient\ Global) * 10\right)\right)}{DAS28} * 100$$

A previous study reported that a DAS28-P >63.1% is the optimal cutoff for identifying the presence of fibromyalgia among patients with RA.[11] The difference between the PtGA and EGA was determined by subtracting the EGA score from the PtGA score.[12]

Statistical analysis. Descriptive statistics were calculated for demographic and clinical characteristics, including composite disease activity measures. Linear regression models were used to examine relationships between measures of disease discordance and QST measures of pain sensitization suggestive of nociplastic pain. The dependent variables were the QST measures, specifically PPT at the trapezius, TS of the forearm, and CPM. For each QST measure, separate unadjusted and adjusted regression models were conducted for each discordance of disease activity measure, with the independent variables being TSJD, DAS28-P, and PtGA – EGA. Age, study site, sex, body mass index, RA symptom duration, seropositivity, depression, and sleep disturbance were included as covariates in the adjusted models.

Sensitivity analyses were conducted to explore the effects of fibromyalgia and systemic inflammation. First, to determine if results were impacted by including individuals with fibromyalgia, regression analyses excluding participants who met the ACR 2016 revised fibromyalgia diagnostic criteria were performed. Second, to examine potential confounding by systemic inflammation, we conducted regression models including CRP level as a covariate. All data analyses were performed using R (version 4.5.0). The alpha level was set at $P < 0.05$ for all analyses.

RESULTS

Participant characteristics. Of the 225 patients included in this study, the average age was 55.6 years old, with an SD of 13.7. Most were White (77.3%), female (81.3%), and seropositive for rheumatoid factor or anti–cyclic citrullinated peptide (71.6%), with an average symptom duration of 12.4 years (SD 12.6) (Table 1). The mean DAS28 was 4.4 ± 1.2. The mean PROMIS depression t score was 50.2 (SD 9.0), whereas the mean PROMIS sleep disturbance score was 54.4 (SD 9.2), both within

Table 1. Characteristics of study participants (n = 225)*

Characteristics	Value
Age, mean ± SD, y	55.6 ± 13.7
Female, n (%)	183 (81.3)
White, n (%)	174 (77.3)
Seropositive, n (%)[a]	161 (71.6)
Body mass index, mean ± SD	28.6 ± 6.9
RA symptom duration, mean ± SD, y	12.4 ± 12.6
Depression, PROMIS t score, mean ± SD	50.2 ± 9.0
Sleep disturbance, PROMIS t score, mean ± SD	54.4 ± 9.2
Fibromyalgia, n (%)	67 (29.8)
Current smoking, n (%)	30 (13.3)
Pain intensity, 0–10 numerical rating scale, mean ± SD	5.2 ± 2.3
Pain interference, PROMIS t score, mean ± SD	60.5 ± 7.3
Oral glucocorticoid use, n (%)	102 (45.3)
Conventional DMARDs, n (%)	92 (40.9)
Biologic/targeted synthetic DMARDs, n (%)	133 (59.1)
Centrally acting pain medications, n (%)	17 (7.6)
Disease Activity Score in 28 joints, mean ± SD	4.4 ± 1.2
Clinical Disease Activity Index, mean ± SD	24.0 ± 14.0
Tender joint count, mean ± SD	10.9 ± 8.6
Swollen joint count, mean ± SD	5.4 ± 5.2
Patient global assessment, mean ± SD	4.2 ± 2.4
Evaluator global assessment, mean ± SD	3.5 ± 2.2
C-reactive protein, mean ± SD, mg/L	8.0 ± 12.5

* DMARD, disease-modifying antirheumatic drug; PROMIS, Patient-Reported Outcomes Measurement Information System; RA, rheumatoid arthritis.
[a] Seropositive for rheumatoid factor and/or anti–cyclic citrullinated peptide.

population normative levels. Median values for PPT at the trapezius, TS of the forearm, and CPM were 2.6 kgf (interquartile range [IQR] 1.8–3.7), 8.3 (IQR 1.7–20.0), and 1.3 (IQR 1.2–1.5), respectively.

The mean TSJD was 5.4, with an SD of 8.2. Eighty-two (36.4%) had a TSJD ≥7 (Supplementary Table 1). The mean DAS28-P was 49.7%, with an SD of 13.3%. Twenty-five (11.1%) had a DAS28-P >63.1%, the previously identified cutoff for fibromyalgia. The mean PtGA - EGA was 0.7 (SD 2.2) (Figure 1).

Associations between measures of disease discordance and QST measures. In the unadjusted linear regression models (Table 2), higher TSJD was significantly associated with lower PPTs at the trapezius (β = –0.06; 95% confidence interval [CI] –0.08 to –0.03), higher TS of the forearm (β = 0.34; 95% CI 0.12 to 0.56), and higher CPM (β = 0.01; 95% CI 0.002 to 0.01). Higher DAS28-P was significantly associated with lower PPTs at the trapezius (β = –0.05; 95% CI –0.07 to –0.04) and higher TS of the forearm (β = 0.25; 95% CI 0.11 to 0.38). PtGA – EGA was not significantly associated with PPT at the trapezius, TS of the forearm, or CPM.

In the adjusted multivariable models (Table 2), higher TSJD remained significantly associated with lower PPTs at the trapezius (β = –0.05; 95% CI –0.08 to –0.02), higher TS of the forearm (β = 0.29; 95% CI 0.05 to 0.53), and higher CPM (β = 0.01; 95% CI 0.002 to 0.02). Higher DAS28-P was significantly associated with lower PPTs at the trapezius (β = –0.05; 95% CI –0.07 to –0.04) and higher TS of the forearm (β = 0.21; 95% CI 0.06 to 0.35). PtGA – EGA was not significantly associated with PPT at the trapezius, TS of the forearm, or CPM.

Sensitivity analyses. In sensitivity analyses excluding participants who met the ACR 2016 revised fibromyalgia diagnostic criteria, higher TSJD remained significantly associated with lower PPTs at the trapezius, higher TS at the forearm, and higher CPM (Supplementary Table 2). We also conducted another sensitivity analysis including CRP level as an additional covariate, and findings were consistent with previous results (Supplementary Table 3).

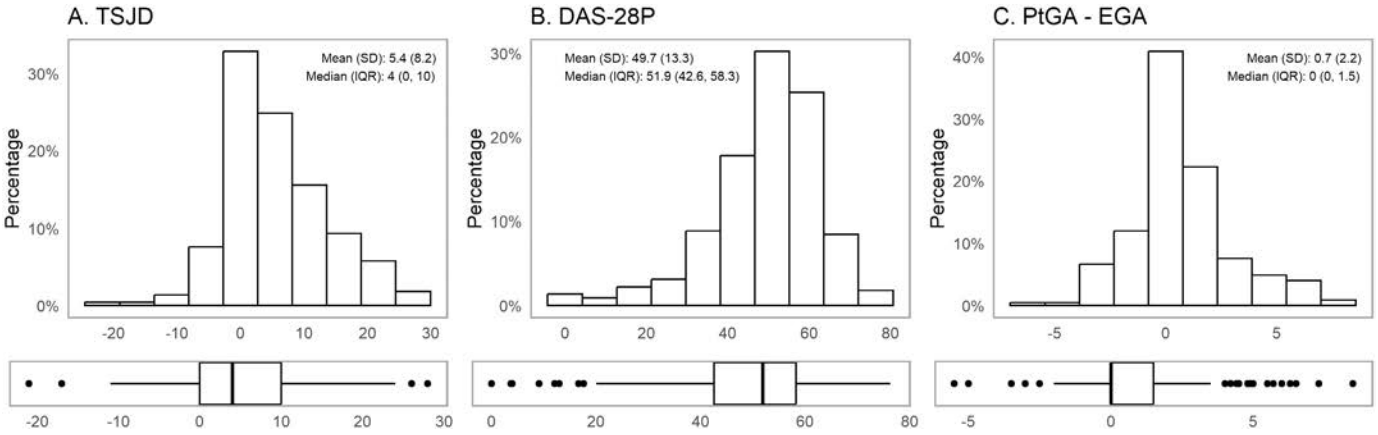


Figure 1. Distribution for disease discordance measures (n = 225). Distribution of disease discordance measures in the Central Pain in Rheumatoid Arthritis cohort, depicted by histogram (with mean and SD) and boxplot (with bold lines indicating median value and edges of boxes corresponding to 25th [lower quartile] and 75th percentiles [upper quartile]). (A) TSJD. (B) DAS28-P. (C) PtGA – EGA. DAS28-P, proportion of subjective components of the Disease Activity Score in 28 joints to the total Disease Activity Score in 28 joints; IQR, interquartile range; PtGA – EGA, patient global assessment of disease activity minus evaluator global assessment of disease activity; TSJD, tender–swollen joint count difference.

Table 2. Associations between discordance measures and QST measures from unadjusted and adjusted linear regression models (n = 225)*

Discordance measures	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Trapezius PPT		
TSJD	**−0.06 (−0.08 to −0.03)**	**−0.05 (−0.08 to −0.02)**
DAS28-P	**−0.05 (−0.07 to −0.04)**	**−0.05 (−0.07 to −0.04)**
PtGA − EGA	0.08 (−0.18 to 0.02)	0.07 (−0.17 to 0.04)
TS of the forearm		
TSJD	**0.34 (0.12 to 0.56)**	**0.29 (0.05 to 0.53)**
DAS28-P	**0.25 (0.11 to 0.38)**	**0.21 (0.06 to 0.35)**
PtGA − EGA	0.30 (−0.55 to 1.1)	0.75 (−0.17 to 1.70)
CPM		
TSJD	**0.01 (0.002 to 0.01)**	**0.01 (0.002 to 0.02)**
DAS28-P	0.003 (−0.0004 to 0.01)	0.003 (−0.001 to 0.01)
PtGA − EGA	0.00 (−0.02 to 0.02)	0.01 (−0.01 to 0.03)

* Models were adjusted for the following covariates: age, study site, sex, body mass index, rheumatoid arthritis symptom duration, seropositivity, depression, and sleep disturbance. Bold values indicate significantly different values ($P < 0.05$). CI, confidence interval, CPM, conditioned pain modulation; DAS28-P, proportion of subjective components of the Disease Activity Score in 28 joints to the total Disease Activity Score in 28 joints; OR, odds ratio; PPT, pain pressure threshold; PtGA − EGA, patient global assessment of disease activity minus evaluator global assessment of disease activity; QST, quantitative sensory testing; TS, temporal summation; TSJD, tender–swollen joint count difference.

DISCUSSION

In this cohort of patients with active RA, greater discordances in certain disease activity metrics were associated with worse QST measures of pain sensitization, suggestive of nociplastic pain. However, the strengths of associations were modest, and not all measures of disease discordance were associated with all QST assessments. Consistent patterns included the following: (1) higher TSJD and DAS28-P were both associated with lower PPT at the trapezius and higher TS, and (2) PtGA – EGA was not significantly associated with any QST measures.

Higher TSJD and DAS28-P were both consistently associated with lower PPT at the trapezius and higher TS. The consistency in these relationships provides confidence in these results, although the strengths of associations were modest. TSJD and DAS28-P are composite measures that incorporate the concept of discordance in tender and swollen joint counts in the assessment of nociplastic pain. Tender joint count is similar to PPT (use of a pressure algometer) and TS (use of punctate probes), in that all assess sensitivity to external applications of pressure. Individuals with higher TS are likely to have greater pain sensitivity (lower PPTs) and more tender joints, leading to higher TSJDs and higher DAS28-P scores.

Conversely, PtGA – EGA was not significantly associated with PPT at the trapezius, TS, or CPM. Neither PtGA nor EGA includes a direct assessment of tender and swollen joint counts. A large observational study previously found EGA to be highly correlated with swollen joint count ($\rho > 0.70$) and PtGA to be the most highly correlated with patient-reported pain intensity ($\rho > 0.86$) and function assessed by the Health Assessment Questionnaire ($\rho > 0.49$).[13] Patient-reported pain intensity and function are important health outcomes that have multiple different causes. Thus, although nociplastic pain may contribute to the discrepancy between PtGA and EGA, there are likely multiple other causes for this discrepancy, such as general health status, mental health, and functional capacity.[28] In addition, previous studies have suggested that patient education levels, patient health literacy, cultural factors, and evaluator experience may also contribute to these discrepancies.[29–32] Lastly, the EGA may be subject to interpretive variability among clinicians. Some assessors may regard it as a measure of inflammatory activity exclusively, in alignment with its application in clinical trial settings, whereas others may integrate a more comprehensive appraisal of disease burden, encompassing pain, fatigue, and functional impairment.[33]

Notably, higher TSJD was weakly associated with higher CPM, but DAS28-P was not associated with CPM. Although PPT at the trapezius is a reflection of overall central nervous system pain modulation, TS and CPM have been thought to reflect specific pain pathways: ascending pain facilitation and descending pain inhibition, respectively.[34] Ascending pain facilitation is commonly thought to be stimulated by peripheral input, whereas descending pain inhibition is primarily thought to be determined by genetics and is less likely to change when peripheral nociceptive input is modulated.[35,36] Given the weak association between TSJD and CPM and the lack of consistency in associations across TSJD and DAS28-P, we caution against making strong inferences regarding the association (or lack thereof) among TSJD, DAS28-P, and CPM. It is possible that, in a population of patients with RA and active peripheral joint inflammation, ascending pain facilitation may be the main contributor to pain sensitivity, whereas descending pain inhibition may play a larger role when peripheral inflammation has been treated. It should be noted that even in the prototypical nociplastic condition, fibromyalgia, these QST test results are only abnormal in a subset of patients.[37,38]

Prior studies have suggested various thresholds for tender–swollen joint discrepancies for nociplastic pain. Pollard et al identified a ≥7 TSJD for detecting patients with fibromyalgia RA, with high sensitivity and specificity.[10] Other studies have suggested that a tender–swollen joint discrepancy as low as 1 may contribute to misrepresentation of RA inflammatory disease activity.[39,40] Additional research is needed to establish a specific threshold for nociplastic pain using tender and swollen joint counts; however, it is reasonable to conclude that the larger the TSJD, the greater the likelihood of nociplastic pain involvement in patients with RA and chronic pain.[41] Between TSJD and DAS28-P, TSJD demonstrates greater feasibility for implementation in a clinical setting because of its ease of basic calculation and assessment. Use of TSJD in clinical practice may identify patients who would benefit from stronger reliance on exercise, psychological

treatment, and noninflammatory pharmacologic treatment and less focus on DMARDs alone for pain management.

Study strengths include the large sample size as well as involvement of multiple centers, which increases the generalizability of our results. To our knowledge, the CPIRA is the largest cohort of patients with RA providing comprehensive data on multiple QST measures, including PPT at the trapezius, TS, and CPM. These QST measures have also been previously demonstrated to be associated with pain intensity and treatment response in patients with RA.[17,42] This study also sought to reduce the heterogeneity of the QST evaluators by use of standardized QST protocol and rigorous training standards.[23]

Limitations of this study include that this was a cross-sectional analysis of patients with active RA, so results are only specific to this disease state. The cross-sectional analysis also precludes causal inferences. Another limitation is that, although QST measures are commonly used to assess nociplastic pain in research studies, there is no gold standard for assessing nociplastic pain, and QST tests are not consistently different in individuals with nociplastic pain (eg, individuals with fibromyalgia) than in healthy controls.[37,38] In addition, a further limitation is the lack of objective and sensitive assessment of synovitis, such as ultrasonography or magnetic resonance imaging, which could have enabled a more accurate characterization of joint inflammation. Moreover, the inclusion of patients taking centrally acting neuropathic pain medications, such as duloxetine or pregabalin, may influence pain processing. However, we could not evaluate this rigorously because only 7.6% of the sample were taking centrally acting medications.

In summary, discordances in disease activity metrics (eg, TSJD and DAS28-P) are useful to consider as potential indicators of nociplastic pain in the clinical setting. By identifying or excluding nociplastic pain as a potential contributor to patient symptoms, health care providers can provide more individualized and targeted recommendations about pain management. This could range from recommending intensification of immunosuppressive therapy to treat inflammatory pain to recommending the addition of a tricyclic antidepressant or serotonin–norepinephrine reuptake inhibitor for nociplastic pain. We would favor TSJD over DAS28-P in clinical practice for identifying these patients, given its ease of calculation. Although our analysis focused on routinely collected clinical measures such as TSJD, DAS28-P, and PtGA – EGA, we acknowledge that other patient-reported measures may present potential practical and reliable alternatives. Future research should investigate the comparative utility of these self-report instruments, given their feasibility and applicability in routine rheumatologic practice.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. ten Klooster PM, Veehof MM, Taal E, et al. Changes in priorities for improvement in patients with rheumatoid arthritis during 1 year of anti-tumour necrosis factor treatment. Ann Rheum Dis 2007;66(11):1485–1490.
2. International Association for the Study of Pain. Terminology. Accessed April 7, 2025. https://www.iasp-pain.org/resources/terminology/
3. Lee YC, Frits ML, Iannaccone CK, et al. Subgrouping of patients with rheumatoid arthritis based on pain, fatigue, inflammation and psychosocial factors. Arthritis Rheumatol 2014;66(8):2006–2014.
4. Schmukler J, Li T, Block JA, et al. RheuMetric physician 0 to 10 estimates of inflammation, damage, and patient distress at initial versus follow-up visits in contemporary rheumatology care. ACR Open Rheumatol 2025;7(3):e70010.
5. Murphy AE, Minhas D, Clauw DJ, et al. Identifying and managing nociplastic pain in individuals with rheumatic diseases: a narrative review. Arthritis Care Res (Hoboken) 2023;75(10):2215–2222.
6. Rolke R, Magerl W, Campbell KA, et al. Quantitative sensory testing: a comprehensive protocol for clinical trials. Eur J Pain 2006;10(1):77–88.
7. Giesecke T, Williams DA, Harris RE, et al. Subgrouping of fibromyalgia patients on the basis of pressure-pain thresholds and psychological factors. Arthritis Rheum 2003;48(10):2916–2922.
8. Rehm S, Sachau J, Hellriegel J, et al. Pain matters for central sensitization: sensory and psychological parameters in patients with fibromyalgia syndrome. Pain Rep 2021;6(1):e901.
9. Greenspan JD, Slade GD, Rathnayaka N, et al. Experimental pain sensitivity in subjects with temporomandibular disorders and multiple other chronic pain conditions: the OPPERA prospective cohort study. J Oral Facial Pain Headache 2020;34(suppl):s43–s56.
10. Pollard LC, Kingsley GH, Choy EH, et al. Fibromyalgic rheumatoid arthritis and disease assessment. Rheumatology (Oxford) 2010;49(5):924–928.
11. McWilliams DF, Zhang W, Mansell JS, et al. Predictors of change in bodily pain in early rheumatoid arthritis: an inception cohort study. Arthritis Care Res (Hoboken) 2012;64(10):1505–1513.
12. Salaffi F, Di Carlo M, Carotti M, et al. The subjective components of the Disease Activity Score 28-joints (DAS28) in rheumatoid arthritis patients and coexisting fibromyalgia. Rheumatol Int 2018;38(10):1911–1918.
13. Studenic P, Radner H, Smolen JS, et al. Discrepancies between patients and physicians in their perceptions of rheumatoid arthritis disease activity. Arthritis Rheum 2012;64(9):2814–2823.
14. Moore MN, Wallace BI, Song J, et al. Correlation of fibromyalgia survey questionnaire and quantitative sensory testing among patients with active rheumatoid arthritis. J Rheumatol 2022;49(9):1052–1057.
15. Wallace B, Moore MN, Heisler AC, et al. 292 fibromyalgianess and glucocorticoid persistence among patients with rheumatoid arthritis. J Clin Transl Sci [abstract] 2022;6(s1):51.
16. Aletaha D, Neogi T, Silman AJ, et al. Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010;69

(9):1580-1588. doi: 10.1136/ard.2010.138461. Erratum in: Ann Rheum Dis. 2010 Oct;69(10):1892.

17. Heisler AC, Song J, Muhammad LN, et al. Association of dysregulated central pain processing and response to disease-modifying antirheumatic drug therapy in rheumatoid arthritis. Arthritis Rheumatol 2020; 72(12):2017–2024.
18. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23(5 suppl 39):S100–S108.
19. Hays RD, Bjorner JB, Revicki DA, et al. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. Qual Life Res 2009;18(7):873–880.
20. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010;150(1):173–182.
21. Trigg A, Haberland C, Shalhoub H, et al. Comparative performance of PROMIS Sleep Disturbance computerized adaptive testing algorithms and static short form in postmenopausal women. J Patient Rep Outcomes 2025;9(1):18.
22. Wolfe F, Clauw DJ, Fitzcharles M-A, et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46(3):319–329.
23. Lee YC, Bingham CO III, Edwards RR, et al. Association between pain sensitization and disease activity in patients with rheumatoid arthritis: a cross-sectional study. Arthritis Care Res (Hoboken) 2018;70(2): 197–204.
24. Suokas AK, Walsh DA, McWilliams DF, et al. Quantitative sensory testing in painful osteoarthritis: a systematic review and meta-analysis. Osteoarthritis Cartilage 2012;20(10):1075–1085.
25. Graven-Nielsen T, Arendt-Nielsen L. Assessment of mechanisms in localized and widespread musculoskeletal pain. Nat Rev Rheumatol 2010;6(10):599–606.
26. Clauw DJ, Arnold LM, McCarberg BH; FibroCollaborative. The science of fibromyalgia. Mayo Clin Proc 2011;86(9):907–911.
27. Pud D, Granovsky Y, Yarnitsky D. The methodology of experimentally induced diffuse noxious inhibitory control (DNIC)-like effect in humans. Pain 2009;144(1–2):16–19.
28. Khan NA, Spencer HJ, Abda E, et al. Determinants of discordance in patients' and physicians' rating of rheumatoid arthritis disease activity. Arthritis Care Res (Hoboken) 2012;64(2):206–214.
29. Nicolau G, Yogui MM, Vallochi TL, et al. Sources of discrepancy in patient and physician global assessments of rheumatoid arthritis disease activity. J Rheumatol 2004;31(7):1293–1296.
30. Hirsh JM, Boyle DJ, Collier DH, et al. Health literacy predicts the discrepancy between patient and provider global assessments of rheumatoid arthritis activity at a public urban rheumatology clinic. J Rheumatol 2010;37(5):961–966.
31. Kaneko Y, Takeuchi T, Cai Z, et al. Determinants of patient's global assessment of disease activity and physician's global assessment of disease activity in patients with rheumatoid arthritis: a post hoc analysis of overall and Japanese results from phase 3 clinical trials. Mod Rheumatol 2018;28(6):960–967.
32. Turk M, Pope JE. Physician global assessments for disease activity in rheumatoid arthritis are all over the map! RMD Open 2018;4(1): e000578.
33. Harrington JT. The uses of disease activity scoring and the physician global assessment of disease activity for managing rheumatoid arthritis in rheumatology practice. J Rheumatol 2009;36(5):925–929.
34. Iyer P, Lee YC. Why it hurts: the mechanisms of pain in rheumatoid arthritis. Rheum Dis Clin North Am 2021;47(2):229–244.
35. Staud R. The important role of CNS facilitation and inhibition for chronic pain. Int J Clin Rheumatol 2013;8(6):639–646.
36. Chen Q, Heinricher MM. Descending control mechanisms and chronic pain. Curr Rheumatol Rep 2019;21(5):13.
37. Schrepf A, Williams DA, Gallop R, et al; MAPP Research Network. Sensory sensitivity and symptom severity represent unique dimensions of chronic pain: a MAPP Research Network study. Pain 2018; 159(10):2002–2011.
38. Harte SE, Ichesco E, Hampson JP, et al. Pharmacologic attenuation of cross-modal sensory augmentation within the chronic pain insula. Pain 2016;157(9):1933–1945.
39. Hammer HB, Michelsen B, Provan SA, et al. Tender joint count and inflammatory activity in patients with established rheumatoid arthritis: results from a longitudinal study. Arthritis Care Res (Hoboken) 2020; 72(1):27–35.
40. Hammer HB, Jensen Hansen IM, Järvinen P, et al. Rheumatoid arthritis patients with predominantly tender joints rarely achieve clinical remission despite being in ultrasound remission. Rheumatol Adv Pract 2021;5(2):rkab030.
41. Eberhard A, Bergman S, Mandl T, et al. Joint tenderness at 3 months follow-up better predicts long-term pain than baseline characteristics in early rheumatoid arthritis patients. Rheumatology (Oxford) 2024; 63(3):734–741.
42. Heisler AC, Song J, Dunlop DD, et al. Association of pain centralization and patient-reported pain in active rheumatoid arthritis. Arthritis Care Res (Hoboken) 2020;72(8):1122–1129.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 54–65
DOI 10.1002/acr.25654

Development and External Validation of a Multivariable Predictive Model for Progression to Difficult-to-Treat Rheumatoid Arthritis in Biologic-Experienced Patients

Misti L. Paudel,[1,2] Nancy Shadick,[1,2] Michael Weinblatt,[1,2] George Reed,[3] Heather J. Litman,[4] Joel M. Kremer,[5] Dimitrios A. Pappas,[4,5,6] and Daniel H. Solomon[1,2]

Objective. Approximately 20% of patients with rheumatoid arthritis (RA) cycle through multiple therapies without achieving treatment goals and are classified as having "difficult-to-treat" RA (D2T-RA); however, no risk prediction tools exist to identify which patients are at highest risk. Our aim was to develop and validate a predictive model for progression to D2T-RA among patients with RA.

Methods. We used data from two large independent observational cohorts of patients with RA to develop and externally validate a multivariable prediction model to identify participants at risk of D2T-RA, defined using EULAR 2021 criteria. We developed a multivariable predictive model for D2T-RA using random survival forests in participants treated with their first biologic and/or targeted synthetic disease-modifying antirheumatic drug (b/tsDMARD) (derivation cohort). We validated the model in a cohort of participants initiating or switching b/tsDMARD therapies.

Results. A total of 700 participants were in the derivation cohort (84% female, mean age 55 years, median follow-up 40 months, 113 [16%] with D2T-RA), and 2,070 participants were included in the validation cohort (79% female, mean age 56 years, median follow-up 8 months, 571 [28%] with D2T-RA). We observed C-index values of 0.643 (95% confidence interval [CI] 0.585–0.698; derivation cohort) and 0.620 (95% CI 0.596–0.643; validation cohort). Calibration measures suggested overall moderate predictive ability. Worsened functional status, pain, fatigue, and global disease activity were consistently top predictors across both cohorts.

Conclusion. Our model demonstrated moderate discrimination and calibration, highlighting the challenge in accurately predicting D2T-RA outcomes. These findings underscore the need for further research to improve predictive performance, potentially through the incorporation of additional biomarkers.

INTRODUCTION

Rheumatoid arthritis (RA) is a chronic inflammatory disease that impacts 0.3% to 1.0% of people worldwide and approximately 1.3 million adults in the United States.[1,2] The past several decades have observed advancements in RA management, in part due to the increasing availability of new biologic and/or targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs).[3] Despite advancements in therapies, a subset of patients with RA remain persistently disabled with chronic pain and functional limitations.

Prior studies have observed that approximately one in five patients with RA cycle through multiple treatments without experiencing relief in symptoms and/or remission and are classified as having "difficult-to-treat" RA (D2T-RA).[4–7] Additionally, one-sixth of patients with RA have been treated with five different classes of b/tsDMARD therapies.[7] Treatment nonresponse and cycling poses a significant challenge for rheumatologists due to

Supported by the NIH (grant P30-AR-072577) and CorEvitas, LLC. The BRASS cohort is additionally supported by Bristol Myers Squibb, Sanofi, Aqtual, and Janssen Pharmaceuticals, Inc. CorEvitas, LLC, has been supported through contracted subscriptions in the last two years by AbbVie, Amgen, Inc, Arena Pharmaceuticals, Boehringer Ingelheim, Bristol Myers Squibb, Chugai, Eli Lilly and Company, Genentech, GSK, Janssen Pharmaceuticals, Inc, LEO Pharma, Novartis, Ortho Dermatologics, Pfizer, Inc, Sun Pharmaceutical Industries Inc, and UCB S.A.

[1]Brigham and Women's Hospital, Boston, Massachusetts; [2]Harvard Medical School, Boston, Massachusetts; [3]University of Massachusetts Chan Medical School, Worcester, Massachusetts; [4]CorEvitas LLC, Waltham, Massachusetts; [5]Corrona Research Foundation, Albany, New York; [6]Columbia University, New York, New York.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25654).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25654.

Address correspondence via email to Misti L. Paudel, PhD, MPH, at mpaudel1@bwh.harvard.edu.

Submitted for publication May 14, 2025; accepted in revised form September 2, 2025.

SIGNIFICANCE & INNOVATIONS

- This is the first study to develop and externally validate a risk prediction model to identify which patients with rheumatoid arthritis (RA) are at highest risk of progressing to "difficult-to-treat" RA (D2T-RA).
- Participants with worse functioning, greater pain, fatigue, and disease activity have increased risk of progressing to D2T-RA.
- These findings underscore the need for further research to improve predictive performance through the incorporation of additional biomarkers.

limited evidence on how to identify and manage patients.[8] In 2021, after reviewing results of an international survey[9] and existing literature, EULAR published criteria for assessing D2T-RA, stating that three conditions must all be present: (1) multiple DMARD failure, (2) signs suggestive of active or progressive disease, and (3) management of symptoms viewed as problematic according to the rheumatologist and/or patient.[10] Some have criticized this definition as too broad,[11] but its strength lies in the requirement of the three domains of treatment cycling, persistent disease activity, and management difficulties, ensuring accurate identification of an important phenotype of this subgroup of patients with RA.

Prior studies have identified several clinical characteristics that are potential risk factors for D2T-RA, including older age, higher rheumatoid factor (RF) levels, higher Clinical Disease Activity Index (CDAI) scores, lower methotrexate dosage, a greater proportion with hypertension, a greater proportion with diabetes, and greater treatment delay.[12] Finally, higher RF levels, greater disease activity, and comorbid pulmonary disease were associated with D2T-RA risk in another cohort.[13] A predictive model that incorporates clinical, demographic, and patient-reported outcomes (PROs) to predict risk of progressing to D2T-RA would be a clinically useful tool to assist rheumatologists in identifying patients with RA who could benefit from alternative treatment and management approaches.

We used data from two longitudinal cohorts of patients with RA with comprehensive longitudinal assessments of RA-related treatments, clinical factors, and PROs to develop and externally validate a predictive model to identify which patients with RA are most likely to advance to this state of multitreatment failure.

PATIENTS AND METHODS

Data sources. We used data from the Brigham and Women's Rheumatoid Arthritis Sequential Study (BRASS) and the Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions (CERTAIN). The BRASS registry is a single-center prospective observational cohort study that enrolled participants from the Brigham and Women's Hospital Arthritis Center in Boston, Massachusetts, from March 2003 to May 2019. Participants were eligible if they had a diagnosis of RA using American College of Rheumatology RA classification criteria[14,15] or based on the opinion of a rheumatologist. Participants completed annual visits with a rheumatologist and interim six-month questionnaires. Study follow-up continued through April 2023. Details of the BRASS registry, including its overall goals and design, have been reported elsewhere.[16] All participants provided informed consent, and approval was obtained by the Institutional Review Board of Brigham and Women's Hospital (2002P001762). The BRASS clinical trial is registered at ClinicalTrials.gov (identifier: NCT01793103).

The CERTAIN study is a 12-month prospective substudy of the large multicenter US-based registry CorEvitas (formerly Corrona). Patients with at least moderate disease activity (CDAI score >10) and those who were starting or switching to a new biologic agent that was an anti–tumor necrosis factor (anti-TNF; adalimumab, etanercept, certolizumab, golimumab, or infliximab) or a non–anti-TNF agent (tocilizumab, abatacept, or rituximab) were eligible to participate. Participants completed visits every three months for up to one year, at which point longitudinal follow-up continued in the parent CorEvitas registry. Participants needing to discontinue and switch to a different biologic agent during CERTAIN follow-up were offered an opportunity to re-enroll in the study, provided they met study inclusion criteria. Details of the CERTAIN study have been reported elsewhere.[17] All participants provided informed consent, and approval was obtained by the New England Institutional Review Board (NEIRB 120160610). The CERTAIN clinical trial is registered at ClinicalTrials.gov (identifier: NCT01625650). For academic investigative sites in CERTAIN that did not receive a waiver to use the central institutional review board, approval was obtained from the governing institutional review boards, and documentation of approval was submitted to the sponsor before any study procedures were initiated. All CERTAIN participants were required to provide written informed consent before participating.

Study design. We used data from BRASS to develop a predictive algorithm for progression to D2T-RA, and we externally validated the algorithm using data from CERTAIN. For brevity, we will refer to these as the derivation and validation cohorts throughout the article. In the BRASS cohort, we restricted to participants with RA who were treated with their first b/tsDMARD and defined the index date as the earliest recorded study visit in which first b/tsDMARD treatment was documented. In CERTAIN, we restricted participants to those who were not currently classified as having D2T-RA at enrollment. Also, for CERTAIN participants who switched therapies and were subsequently re-enrolled, we strung consecutive enrollments together and recalculated follow-up time from study entry. Across both cohorts, patients missing demographic data were excluded. In secondary analyses using

CERTAIN, we created a subset of participants who were treated with their first b/tsDMARD (naïve). Both sets of analyses are presented.

D2T-RA. We mapped available study variables to each component of EULAR's D2T-RA criteria and identified the earliest study visit in which participants met all three components: DMARD failure, persistent disease activity, and perception of a persistent problem during follow-up. Details of EULAR's D2T-RA algorithm and mapped variables for each cohort are provided in Table 1. Details of how this algorithm was implemented in BRASS have been previously published.[18] We made modifications to D2T-RA subcriteria when needed based on the availability of data elements in CERTAIN. These modifications included extra-articular manifestations, prednisone use, and erosions and have been described in the Methods section.

Potential predictors. We included several potential predictors in random survival forests (RSF) models, spanning RA-related treatments, clinical factors, demographics, and PROs. We provide more details of each of these measures from both cohorts in the following paragraphs.

RA-related treatments. RA-related treatments included conventional synthetic DMARD and b/tsDMARD therapies and glucocorticoids. We classified b/tsDMARD use as TNF inhibitors (etanercept, adalimumab, infliximab, golimumab, certolizumab pegol), anti–T cell costimulator (abatacept), anti–interleukin-1 (anti–IL-1) inhibitors (anakinra), B cell depletion (rituximab), anti–IL-6 inhibitors (tocilizumab, sarilumab), and JAK inhibitors (tofacitinib, baricitinib, upadacitinib). Use of glucocorticoids, such as prednisone or methylprednisolone, was assessed at each visit and expressed as the most common daily dosage in the past six months (BRASS) or current treatment (CERTAIN).

Measures of disease activity. Rheumatologists completed the four-variable Disease Activity Score in 28 joints using the C-reactive protein level (DAS28-CRP),[19] which included a tender joint count (TJC) and swollen joint count (SJC), patient global assessment (PtGA), and CRP laboratory values. A CDAI was assessed, with scores >10 indicative of active RA with moderate disease. The CDAI consists of SJC, TJC, PtGA, and provider global assessment (PrGA). For both cohorts, the PtGA scores ranged from zero (arthritis was not at all active) to 100 (arthritis was extremely active).[20,21] Scores ≥30 were indicative of patient perception of active RA disease. In addition, rheumatologists completed the PrGA, with scores ranging from 0 (no arthritis activity) to 100 (severe arthritis activity).[22] Scores ≥30 were indicative of provider perception of active RA disease.

Radiographic measures and erosions. In BRASS, bilateral hand and wrist radiographs were performed at enrollment and every two years during follow-up and were scored by four

Table 1. Description of EULAR criteria for difficult-to-treat RA in BRASS and CERTAIN*

Criterion	EULAR definition	BRASS definition	CERTAIN definition
DMARD failure	Criterion 1: failure of ≥2 b/tsDMARDs with different mechanisms of action after failing conventional synthetic DMARD therapy	Patient-reported treatment with ≥2 b/tsDMARDs with different mechanisms of action for at least 3 mo	Patient-reported treatment with 2 b/tsDMARDs with different mechanisms of action for at least 3 mo
Persistent disease activity	Criterion 2: at least one of the following conditions (a–e)		
	(a) At least moderate disease activity, such as DAS28-ESR >3.2 or CDAI >10	DAS28-CRP >3.2 or CDAI >10	DAS28-CRP >3.2 or CDAI >10
	(b) Signs (acute-phase reactants and imaging) and/or symptoms suggestive of active disease	Presence of extra-articular manifestations in the past year	History of selected extra-articular manifestations
	(c) Inability to taper glucocorticoid treatment below 7.5 mg/day of prednisone or equivalent	Glucocorticoid most frequent average dose in past six months ≥6 mg	Current treatment with prednisone with a reported dose ≥7.5 mg
	(d) Rapid radiographic progression (with or without signs of active disease)	Change in Sharp score of 5 or more points	History of ever having erosive disease
	(e) Well-controlled disease according to above standards but still having RA symptoms that are causing a reduction in quality of life	RAPID3 score >3 (moderate/high disease severity) and does not meet criteria a–d	RAPID3 score >3 (moderate/high disease severity) and does not meet criteria a–d
Perception of a persistent problem	Criterion 3: management of RA and signs/symptoms are perceived as problematic by the rheumatologist and/or the patient	PtGA score >3 or PrGA score >3	PtGA score >3 or PrGA score >3

* BRASS, Brigham and Women's Rheumatoid Arthritis Sequential Study; b/tsDMARD, biologic and/or targeted synthetic DMARD; CDAI, Clinical Disease Activity Index; CERTAIN, Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions; CRP, C-reactive protein; DAS28, Disease Activity Score in 28 joints; DMARD, disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; PtGA, patient global disease activity; PrGA, provider global assessment; RA, rheumatoid arthritis; RAPID3, Routine Assessment of Patient Index Data 3.

Brigham and Women's Hospital radiologists according to the Sharp/van der Heijde method.[23,24] Erosions were scored for 16 joints on each side of the body, and a total erosion score summed erosions in hands and wrists. Joint space narrowing (JSN) was summed across hands and wrists,[24,25] ranging from 0 to 120. The Sharp/van der Heijde score was computed as the total erosion score plus the total JSN score.[26] In CERTAIN, a history of joint erosions was recorded.

Extra-articular manifestations. Extra-articular manifestations of RA experienced by participants were recorded in BRASS and included pericarditis, pleuritis, pulmonary fibrosis, bronchiectasis, pulmonary nodules, bronchiolitis obliterans with organizing pneumonia, subcutaneous rheumatoid nodules, Sjögren disease or keratoconjunctivitis sicca, large granular lymphocytic leukemia, cervical myelopathy, neuropathy, Felty syndrome, cutaneous vasculitis, pyoderma gangrenosum, scleritis or episcleritis, glomerulonephritis, vasculitis of other organs, amyloidosis, and lymphadenopathy. In CERTAIN, these consisted of secondary Sjögren disease, history of subcutaneous nodules, interstitial lung disease, pulmonary fibrosis, rheumatoid nodules, and rheumatoid pleurisy.

The Routine Assessment of Patient Index Data 3 (RAPID3)[27] is a patient-reported measure of RA severity that includes three subcomponents on the Multidimensional Health Assessment Questionnaire (MDHAQ)[28] for function, pain, and global estimate of status. Standard cutoffs of disease activity on the RAPID3 were remission (≤3.1), low disease activity (3.1–6), moderate disease activity (6.1–12), and high disease activity (>12).

Demographics and other non–RA-related clinical factors. Demographics (age, sex, education, ethnicity), RA disease duration, comorbid medical conditions (physician-reported history of diabetes mellitus, hypertension, and cardiovascular disease [angina, heart attack, and heart failure]), age at RA onset, and RF and anti–cyclic citrullinated peptide positivity were assessed at enrollment. Body mass index (BMI) and smoking status were assessed at baseline and follow-up, and we used a last value carry forward approach to fill in missing values at interim six-month and missed visits.

Statistical analysis. Descriptive analyses were performed to evaluate demographic and clinical characteristics and PROs in each cohort separately. Follow-up time was computed from the index date (as defined for each cohort) until the earliest visit in which D2T-RA criteria were met or the last study visit. This study was conducted in accordance with the TRIPOD+AI guidelines for transparent reporting of multivariable prediction models using machine learning.[29]

Predictive model. To build a predictive model, we first performed necessary data cleaning and harmonization steps, evaluating outliers and missing values. We excluded variables with very low or high prevalence that are unlikely to have predictive utility (ie, <3% and >97%). We imputed missing values using the impute-only feature in RSF under the assumption that data are missing at random.[30] We used RSF to develop prediction models using the "RandomForestSRC" package in R.[30,31] RSF is an extension of random forest methods for right-censored time-to-event data in which survival trees are generated using bootstrapped data and features (predictors) are selected randomly when splitting tree nodes. RSF constructs predictions by drawing a set of bootstrap samples from the data for the number of specified trees (*ntree*), leaving some observations (approximately one-third) out of bag (OOB). We computed prediction accuracy and variable importance measures on the OOB samples. We tuned the model using an iterative process to optimize *ntree*, *mtry*, and *nodesize* hyperparameters, with the overall goal to minimize the OOB error rate. We evaluated Harrell's concordance index (C-index) from the model, which represents an extension of the area under the receiver operating characteristic curve for survival models and estimates the probability that, for a randomly selected pair of patients, the one who developed D2T-RA earlier had a worst predicted outcome (C-index values of 1 indicate perfect discrimination). We used bootstrapping to obtain 95% confidence intervals (CIs) for C-index values. We calculated variable importance (VIMP) values, which compare model performance with and without the variable included. Initial models were run on 45 total predictors (Supplementary Table S1). After reviewing initial model performance metrics, we removed predictors with VIMP values that were zero or negative (indicating the variable may be negatively impacting model performance) from the model.

We plotted Kaplan–Meier (KM)–like survival curves based on predicted probabilities for each time point from the RSF model. These curves resemble KM plots in appearance and interpretation but differ, as they are created from a machine learning survival model and not a nonparametric KM estimator.

In the RSF model, "calibration" refers to how closely survival times predicted by the model match observed survival times. We plotted calibration curves that evaluate the percentage of prediction error by time, including a reference model curve (null model with no predictors) and a curve for the final predictive model. Values closer to zero indicate better calibration. In addition, we evaluated standardized continuous ranked probability scores (CRPS), which quantify the difference between the predicted cumulative probability distribution and observed outcomes standardized by follow-up time. Values from a standardized CRPS are scaled to range from 0 to 1, with those closer to 0 indicating better probabilistic prediction. Although there is no threshold to indicate adequacy, in general, values below 0.5 suggest substantial improvement over a reference model.

We evaluated overall model performance by assessing the Integrated Brier Score (IBS). The IBS is the average of Brier Scores across each time interval, which are computed by calculating mean squared difference between predicted and actual outcomes for specific time points. Values for IBS range from zero

Table 2. Patient characteristics at baseline in BRASS and CERTAIN*

Characteristic	Derivation cohort (BRASS; N = 700)	Validation cohort (CERTAIN)	
		Full sample (N = 2,070)	b/tsDMARD naïve (n = 1,043)
Follow-up, mo	80 (63)	8.2 (5.7)	9.5 (5.6)
Demographics and comorbidities			
Current age, y	55 (14)	56 (13)	56 (13)
18–44 y	167 (24%)	393 (19%)	202 (19%)
45–65 y	370 (53%)	1,160 (56%)	597 (57%)
≥ 65 y	163 (23%)	517 (25%)	244 (23%)
Age at RA diagnosis, y	42 (15)	49 (14)	51 (14)
Female sex	584 (83%)	1,644 (79%)	817 (78%)
Education			
High school diploma or less	120 (17%)	815 (39%)	435 (42%)
At least some technical college or professional school	154 (22%)		
Graduated college	190 (27%)	1,255 (61%)	608 (58%)
Graduate school	236 (34%)		
Smoking status			
Never smoked	425 (61%)	997 (48%)	513 (49%)
Former smoker	231 (33%)	665 (32%)	305 (29%)
Current smoker	44 (6.3%)	408 (20%)	225 (22%)
Body mass index			
<30	527 (75%)	1,199 (58%)	607 (58%)
≥30	173 (25%)	871 (42%)	436 (42%)
History of type 2 diabetes	36 (5.1%)	174 (8.4%)	87 (8.3%)
History of hypertension	197 (28%)	565 (27%)	271 (26%)
History of cardiovascular disease[a]	52 (7.4%)	1,267 (61%)	616 (59%)
RA-related clinical factors			
Rheumatoid factor positivity	482 (69%)	1,376 (66%)	685 (66%)
Anti-CCP positivity	489 (70%)	1,268 (61%)	616 (59%)
DAS28-CRP	3.4 (1.5)	4.8 (1.1)	4.8 (1.1)
≤3.2	348 (50%)	142 (6.9%)	62 (5.9%)
3.3–5.1	249 (36%)	1,159 (56%)	600 (58%)
>5.1	103 (15%)	769 (37%)	381 (37%)
CDAI >10	404 (58%)	2,046 (99%)	1,031 (99%)
CRP	6.1 (13.7)	10 (18)	11 (19)
Tender joint count (0–28)	6 (7)	11 (7)	11 (7)
Swollen joint count (0–28)	6 (6)	8 (6)	8 (6)
PtGA (VAS 0–100)	4.4 (2.8)	5.3 (2.5)	5.2 (2.5)
PtGA ≥30	486 (69%)	1,773 (86%)	889 (85%)
PrGA (VAS 0–100)	2.7 (1.9)	5.0 (2.0)	5.1 (1.9)
PrGA ≥30	347 (50%)	1,919 (93%)	985 (94%)
RAPID3 >3	479 (68%)	1,960 (95%)	982 (94%)
One or more extra-articular manifestations	156 (22%)	578 (28%)	252 (24%)
MDHAQ total score (0–100 scale)	18 (16)	32 (23)	31 (22)
MDHAQ fatigue scale (0–100)	39 (29)	54 (29)	53 (30)
MDHAQ pain scale (0–100)	30 (25)	54 (27)	53 (27)
Methotrexate concomitant therapy	367 (55%)	1,257 (61%)	710 (68%)
Current prednisone use	139 (21%)	691 (33%)	333 (32%)
b/tsDMARD use			
Adalimumab	177 (25%)	439 (21%)	322 (32%)
Etanercept	409 (58%)	346 (17%)	270 (26%)
Other TNFi	75 (11%)	641 (31%)	309 (30%)
Other b/tsDMARDs	39 (5.6%)	644 (31%)	142 (14%)
b/tsDMARD treatment naïve	700 (100%)	1,043 (50%)	1,043 (100%)

* Values are n (%) or mean (SD). anti-CCP, anti–cyclic citrullinated peptide; BRASS, Brigham and Women's Rheumatoid Arthritis Sequential Study; b/tsDMARD, biologic and/or targeted synthetic disease-modifying antirheumatic drug; CDAI, Clinical Disease Activity Index; CERTAIN, Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions; DAS28-CRP, Disease Activity Score in 28 joints using the C-reactive protein level; MDHAQ, Multidimensional Health Assessment Questionnaire; PrGA, provider global assessment; PtGA, patient global assessment; RA, rheumatoid arthritis; RAPID3, Routine Assessment of Patient Index Data 3; TNFi, tumor necrosis factor inhibitor; VAS, visual analog scale.

[a] Includes angina, heart attack, and heart failure.

(perfect prediction) to upward. There are no thresholds to indicate adequacy; however, in general, lower IBS values indicate overall better predictive performance, and values for models with good performance often range between 0.1 and 0.3, although this can vary. We evaluated Shapley Additive Explanations (SHAP)[32] values to better understand the contribution of each variable in the model to the overall prediction. We present SHAP values in a beeswarm plot, where each row represents a predictor in the model, ranked (highest to lowest) by mean absolute SHAP value. Dots in a beeswarm plot represent the SHAP values for each participant and predictor in the analysis. The color of the dots denotes feature value, with higher values in orange and lower values in dark red. Predictors with a positive SHAP value positively impact the model's prediction of higher risk, whereas those with negative values have a negative impact on the model's prediction (lower risk), and the magnitude of the SHAP value indicates how important that variable was to the prediction.

We externally validated our predictive model in CERTAIN data, assessing the same metrics for discrimination and calibration. In primary analyses, we assessed model performance on the full analytic sample in CERTAIN consisting of b/tsDMARD-naïve and experienced participants. In secondary analyses, we restricted to b/tsDMARD-naïve participants to better align with the characteristics of the derivation cohort, who were all treated with their first b/tsDMARD. We also performed an additional secondary analysis refitting the model in the full sample of CERTAIN participants.

RESULTS

A total of 1,581 participants were enrolled in BRASS, and of these, 708 (45%) had at least one visit in which they were recorded as being treated with their first b/tsDMARD. After excluding 8 (1%) participants with missing demographics, the final derivation cohort consisted of 700 (99%) participants. In the validation cohort, 2,350 participants enrolled in CERTAIN, of whom we excluded 2 (<1%) who had missing demographics and 278 (12%) who met D2T-RA criteria at study entry. Our final analytic sample for validation consisted of 2,070 (88%) participants, of whom 1,043 (50%) were b/tsDMARD naïve (Supplementary Figure S1).

Descriptive information on demographics and RA-related clinical factors and PROs in each cohort are reported in Table 2. In general, the derivation and validation cohorts had similar mean ages (55–56 years) and similar distributions of female sex (83%–78%). We noted differences in BMI, prevalence of cardiovascular disease, and multiple measures of disease activity and PRO measures between the derivation and validation cohorts.

In the derivation cohort, 113 (16%) participants progressed to D2T-RA over a mean of 80 months of follow-up. In initial RSF models, we included 45 variables and calculated VIMP values. We removed 22 variables that had zero or negative VIMP values to improve model performance. The final predictive model consisted of 23 variables, including D2T-RA flag, follow-up time, age, age at RA diagnosis, education, BMI categories, smoking status, history of type 2 diabetes, history of hypertension, history of cardiovascular disease, current prednisone use, DAS28-CRP score, DAS28-CRP categories, CRP level, PrGA score, PrGA categories, PtGA score, SJC, MDHAQ total score, MDHAQ anxiety scale, MDHAQ fatigue scale, MDHAQ pain scale, and RAPID3 score. An overview of all variables considered and included in final models is presented in Supplementary Table S1. After performing model tuning processes, the optimal settings for the final RSF model were 1,000 trees, a terminal node size of 10, and *mtry* of 6. We additionally selected the default of sampling without replacement for model training.

In the final derivation model, we observed a Harrell's C-index of 0.64 (95% CI 0.59–0.70), standardized CRPS of 0.136, and IBS of

Table 3. Model performance metrics for random survival forests*

Metrics	Derivation cohort (BRASS)	Validation cohort (CERTAIN)	
		Full sample	b/tsDMARD-naïve subset
Sample size (*ntree*)	700	2,070	1,043
Number of outcomes (D2T-RA)	113	571	66
Number of trees	1,000	1,000	1,000
Terminal node size (*nodesize*)	10	10	10
Average number of terminal nodes	38.5	38.5	38.5
Total number of variables	23	23	23
Number of variables tried at each split (*mtry*)	6	6	6
CRPS	29.4	85.14	51.7
Standardized CRPS	0.136	0.394	0.239
Integrated Brier Score (full follow-up)	0.064	0.473	0.313
Harrell's C-index (95% CI)	0.64 (0.59–0.70)	0.62 (0.60–0.65)	0.63 (0.54–0.71)

* Sampling without replacement and log rank random splitting were used to grow trees. BRASS, Brigham and Women's Rheumatoid Arthritis Sequential Study; b/tsDMARD, biologic and/or targeted synthetic disease-modifying antirheumatic drug; CERTAIN, Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions; CI, confidence interval; CRPS, continuous ranked probability scores; D2T-RA, difficult-to-treat rheumatoid arthritis.

0.064 (Table 3). Our visual inspection of a prediction error curve by time indicates an overall low prediction error that only slightly increases over time, remaining below a maximum of 0.2 (Figure 1).

In the validation cohort of the full sample, we observed that 571 (28%) participants progressed to D2T-RA over a median of 8.2 months of follow-up. We externally validated the final predictive model in CERTAIN data and obtained a Harrell's C-index of 0.62 (95% CI 0.60–0.65), standardized CRPS of 0.394, and IBS of 0.473 (Table 3). A plot of the prediction error curve by time indicates a nearly linear increase in prediction error across time, with

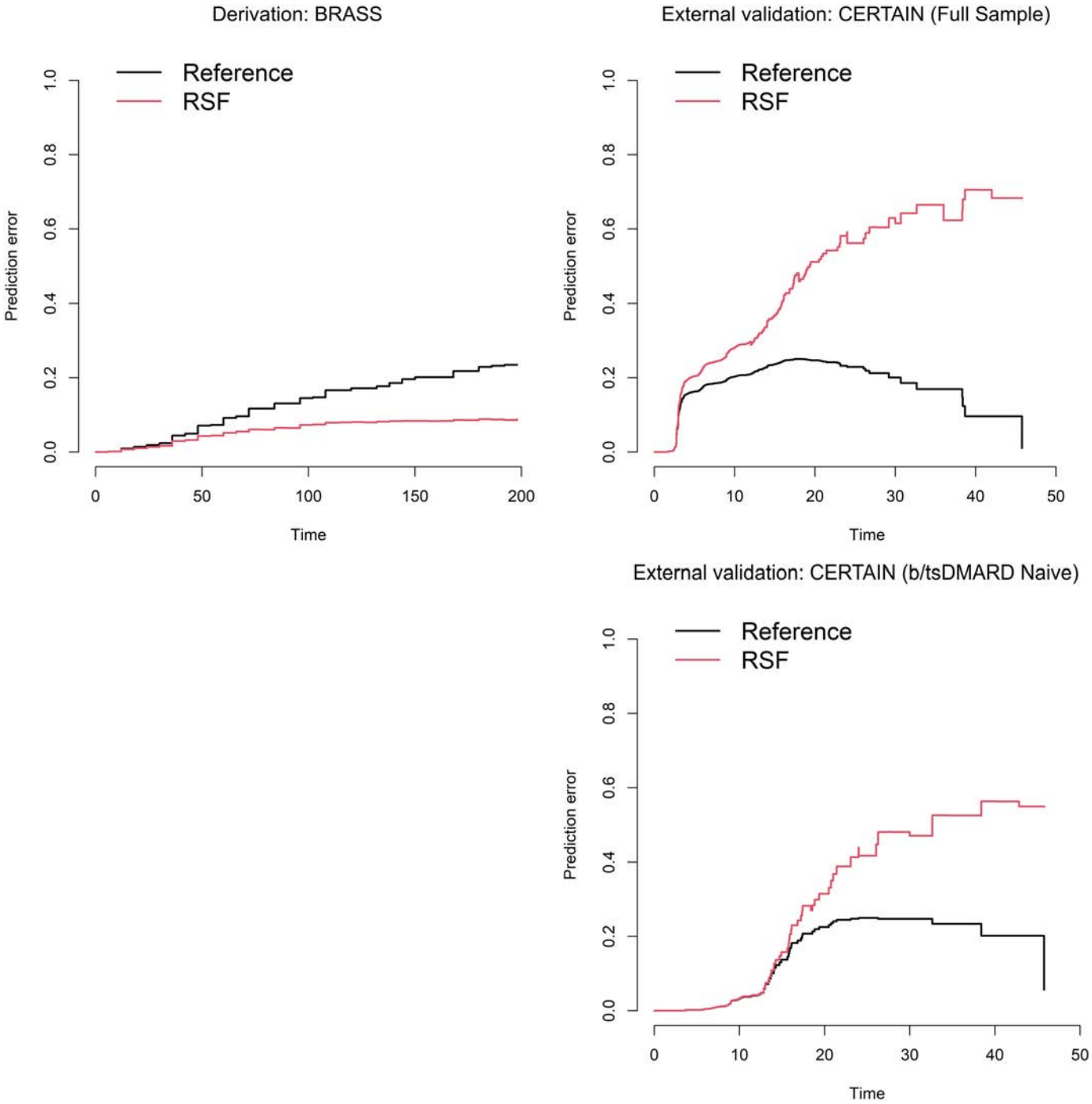


Figure 1. Plot of prediction error by time in BRASS and CERTAIN. These plots depict how prediction error changes over follow-up in the models. The black reference line represents a null model (without predictors), and the red line represents RSF prediction model. Plots are provided for the derivation cohort (BRASS) as well as the validation cohort in the full sample and in the b/tsDMARD-naïve subset. In an ideal scenario, the prediction model would maintain a low, consistent prediction error over time and demonstrate improvement over the reference model. BRASS, Brigham and Women's Rheumatoid Arthritis Sequential Study; b/tsDMARD, biologic and/or targeted synthetic disease-modifying antirheumatic drug; CERTAIN, Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions; RSF, random survival forests.

better calibration observed across shorter follow-up times (Figure 1). In a secondary analysis, we validated the model in a b/tsDMARD-naïve subset of participants enrolled in CERTAIN. We observed that 66 (6%) progressed to D2T-RA over a median of 9.5 months of follow-up. We observed a Harrell's C-index of 0.63 (95% CI 0.54–0.71), standardized CRPS of 0.239, and IBS of 0.313 (Table 3). A visual inspection of the calibration curve suggested improvement in prediction error

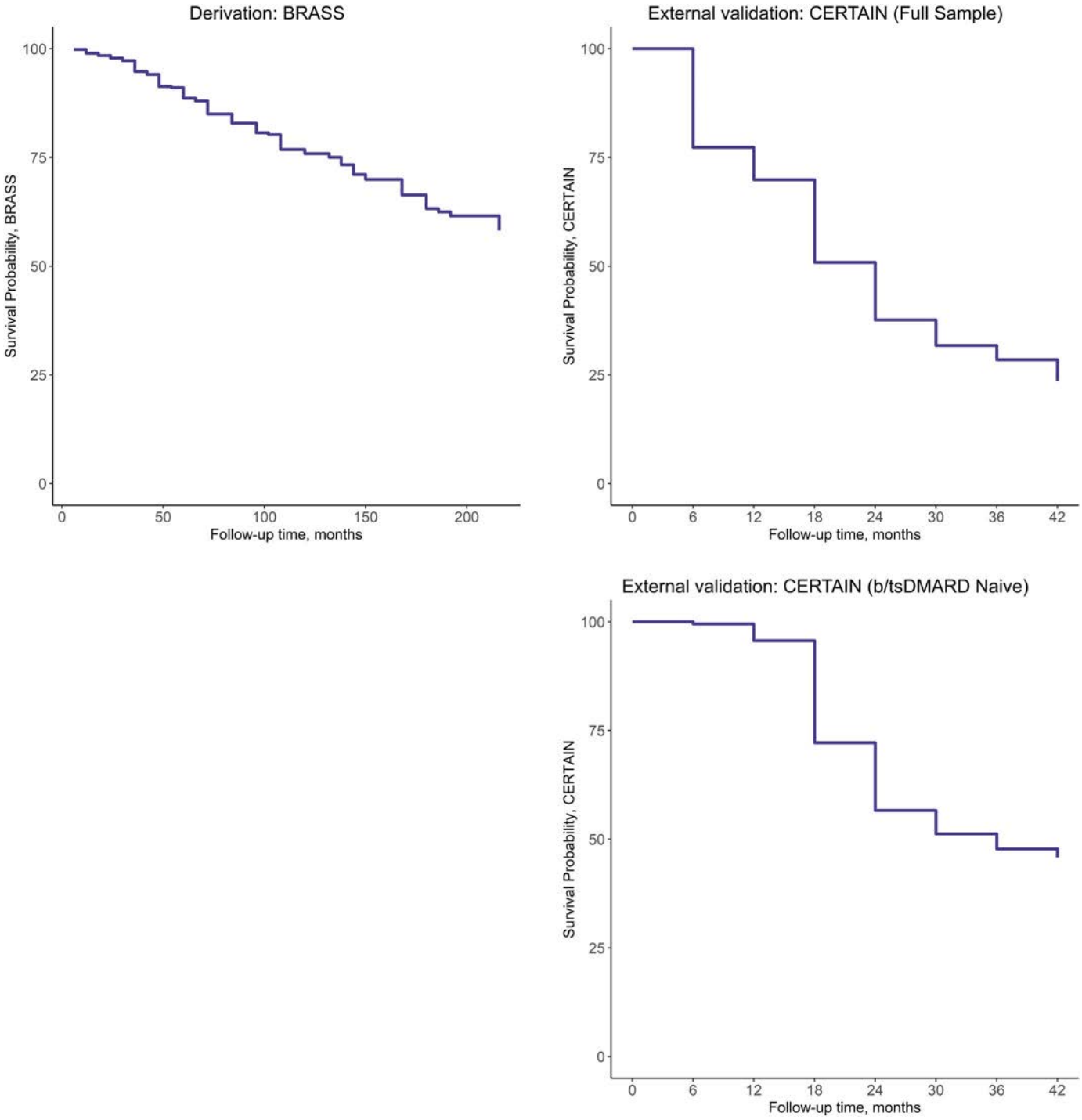


Figure 2. Kaplan–Meier-like survival curves in the derivation and validation cohorts. These curves resemble a Kaplan–Meier plot in appearance and interpretation but were constructed by obtaining and averaging the predicted probabilities for each time point from a machine learning survival model and not a nonparametric Kaplan–Meier estimator. BRASS, Brigham and Women's Rheumatoid Arthritis Sequential Study; b/tsDMARD, biologic and/or targeted synthetic disease-modifying antirheumatic drug; CERTAIN, Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25654/abstract.

for the first 12 to 15 months of follow-up, compared to the full sample (Figure 1).

We present KM-like survival curves in Figure 2 for the derivation and validation cohorts that demonstrate a much steeper decline for participants in the validation cohorts than in derivation cohort, despite differences in follow-up. In our derivation model, we observed that PROs for MDHAQ and PtGA scores were the top predictors in terms of absolute mean SHAP value (Figure 3). In general, feature value was higher for positive SHAP values, indicating that greater values of the MDHAQ or PtGA score were predictive of a higher risk of D2T-RA and did not appear to be linear across most variables. The importance and spread of top predictors in our validation analysis indicated some differences. As an example, the topmost important predictors in validation data were the PtGA score, the MDHAQ pain subscale score, BMI >30, and the RAPID3 score. Additionally, in validation data, smoking status, BMI, and the RAPID3 score appeared to be more important than they were observed to be in the derivation analysis. Finally, results of sensitivity analyses refitting our model to CERTAIN data demonstrated similar discrimination (C-index = 0.62, 95% CI 0.60–0.65; Supplementary Table S2) and calibration to the derivation cohort (CRPS = 0.200; IBS = 0.084).

DISCUSSION

We used data from two large prospective cohorts of individuals with RA to develop and externally validate a predictive model to identify patients at the highest risk of progressing to D2T-RA. Overall, our predictive model demonstrated moderate discrimination and calibration in the derivation and validation cohorts. In addition, our results indicated that top predictors of D2T-RA consistently included PROs of functioning, pain, and fatigue and PtGA scores, whereas additional predictors differed in ranking of VIMP between the two models.

Our finding that PROs of pain, fatigue, and functioning and PtGA scores were top predictors of progression to D2T-RA has not been previously reported. However, a connection between persistent unresolved pain in patients with RA treated with b/tsDMARDs has been reported in the literature. In this prior study, it was noted that inflammation was not able to fully explain pain trajectories and that noninflammatory predictors, such as greater disability and smoking history, were predictive of persistent pain.[33] These findings highlight the importance of integrating PROs into predictive models and the potential clinical value of including them in treatment-related decision-making, as they may reflect unmeasured disease mechanisms such as central sensitization or psychosocial burden, which influence disease progression and treatment response. Future research should continue to explore trajectories of PROs and risk of D2T-RA.

In addition to PROs, we identified several additional predictors of D2T-RA, which included the PrGA score, obesity, and SJC. These results have some overlap but are not in full alignment with prior studies evaluating risk factors for D2T-RA. For example, in a study of 672 patients with RA, investigators observed that over a mean follow-up of 48 months, 8% of participants progressed to D2T-RA, and the primary predictors in a multivariable adjusted logistic regression model were high baseline RF (>156.4 IU/mL), greater DAS28 using the erythrocyte sedimentation rate (ESR), and coexisting pulmonary disease.[12] In our study, RF and extra-articular manifestations, such as pulmonary disease, were not strong predictors of D2T-RA and did not get selected for our final model. However, we similarly observed that the DAS28-CRP score was a top risk factor in both cohorts. Other studies have also reported that female sex, delay in initial treatment, higher disease activity, and/or number of prior treatments are associated with risk of refractory RA.[34,35] We did not observe these same associations, as sex was not selected in our final model based on low VIMP. We did not have a measure for treatment delay in our cohort and were unable to assess this risk factor. Finally, our derivation cohort consisted of participants with RA who were treated with their first b/tsDMARD therapy, and thus the number of prior b/tsDMARDs was not a factor.

We included conventional DMARD therapies in the derivation model and observed that concurrent methotrexate use (meaning in combination with b/tsDMARD) was inversely associated with D2T-RA risk, indicating that participants who were not concurrently treated with methotrexate at the index date had a higher risk of D2T-RA. However, in the validation cohort, concurrent methotrexate use was not identified as an important risk factor, despite it being slightly more prevalent. We are unable to assess why methotrexate nonuse was an important risk factor for D2T-RA, and this should be explored in future research.

Our derivation and validation cohorts are robust longitudinal studies of patients with RA, enriched with several clinical, laboratory, and PRO assessments. They also have several key differences in study design and participant characteristics. Despite these notable differences, our results indicated moderate discrimination, which was similar across both the derivation and validation cohorts. Our results also depicted slightly worsened calibration in the validation cohort, which was improved slightly when we restricted analyses to participants who were b/tsDMARD naïve. These results indicate that our model effectively distinguished between higher- and lower-risk participants but was less accurate with individual-level predictions. However, in the validation cohort, calibration was strongest in the first 12 to 15 months of follow-up, with adequate performance, likely reflecting the greater number of observed events during this period. Taken together, these results highlight the robustness and generalizability of our model in handling variations in patient characteristics. Additionally, the model's ability to generalize to a higher-risk population in our validation (CERTAIN) cohort enhances its clinical utility. It is important to note that our goal was to develop and validate an algorithm to predict the broad definition of D2T-RA that EULAR intended, which captures a heterogenous mix of patients

Figure 3. Beeswarm plot of SHAP values. These plots provide a global overview of the contribution of each of the top 15 predictors in the models. Predictors (features) are ordered vertically from highest to lowest absolute mean SHAP value, and each dot in the rows represents an individual's SHAP value in the dataset. Feature value is indicated by color, with higher values represented by yellow colors. BMI, body mass index; BRASS, Brigham and Women's Rheumatoid Arthritis Sequential Study; b/tsDMARD, biologic and/or targeted synthetic disease-modifying antirheumatic drug; CERTAIN, Comparative Effectiveness Registry to Study Therapies for Arthritis and Inflammatory Conditions; CRP, C-reactive protein; DAS28, Disease Activity Score in 28 joints; GT, greater than; MDHAQ, Multidimensional Health Assessment Questionnaire; Phys, physician; Pt, patient; RAdx, rheumatoid arthritis diagnosis; RAPID3, Routine Assessment of Patient Index Data 3; SHAP, Shapley Additive Explanations; SJC, swollen joint count. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25654/abstract.

with RA who have inflammatory and noninflammatory evidence. It is possible that focusing on a narrower definition of D2T-RA, such as inflammatory D2T-RA, could improve model performance. We were unable to assess this, as we observed very little noninflammatory D2T-RA in prior analyses using BRASS.[18] Therefore, it is important for future studies to continue to validate the model in additional cohorts, including populations with differences in characteristics and/or additional heterogeneity. Navigating heterogeneity across studies highlights a common challenge with predictive modeling and the importance of training models on diverse populations.

This study has many strengths. We used rigorous statistical methodology on two robust longitudinal cohorts of participants with established RA to develop and validate a predictive model in adherence to TRIPOD+AI best practices. We also used a comprehensive definition of D2T-RA based on EULAR criteria and incorporated an extensive set of commonly available clinical and PRO risk factors in our model.

Some limitations we considered are noted herein. We harmonized measures across cohorts to the extent possible; however, some variables were not measured equivalently. We have noted these examples in our Methods section, which include extra-articular manifestations, prednisone use, the MDHAQ anxiety subscale, and erosions. In addition, ESR was not assessed in either of the cohorts, and thus we used the CRP level in our measures of disease activity. The cohorts also had notable differences in clinical characteristics, including length of follow-up and disease activity. As we have noted, differences in characteristics and follow-up between the derivation and validation sets did not have a substantive impact on model performance and calibration. Finally, we used RSF to build predictive models, but we could have selected alternative methods, such as gradient-boosted trees or extremely randomized survival trees. Given that RSF is a robust method, and our data lack high dimensionality, it is unlikely that results would be substantially improved with alternative models.

In summary, we developed and externally validated a predictive model to identify patients with RA who are b/tsDMARD experienced and are at highest risk of progressing to a multitreatment failure state of D2T-RA. Our model demonstrated moderate performance on measures of demographics, RA-related clinical factors, and PROs of pain, fatigue, and functioning. Future research should continue to better understand the role of pain, fatigue, and functioning in multitreatment failure and explore additional predictors, such as candidate inflammatory biomarkers, to improve model performance and assess clinical utility.

ACKNOWLEDGMENTS

The authors would like to thank all the participating investigators, study staff, and participants in the BRASS and CorEvitas RA registries who contributed data for this study.

AUTHOR CONTRIBUTIONS

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

ROLE OF THE STUDY SPONSOR

CorEvitas, LLC employees were involved in the study design, collection, interpretation of the data, the writing of the manuscript, and the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by CorEvitas, LLC.

REFERENCES

1. Almutairi K, Nossent J, Preen D, et al. The global prevalence of rheumatoid arthritis: a meta-analysis based on a systematic review. Rheumatol Int 2021;41(5):863–877.
2. Hunter TM, Boytsov NN, Zhang X, et al. Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004-2014. Rheumatol Int 2017;37(9):1551–1557.
3. Fraenkel L, Bathon JM, England BR, et al. 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2021;73(7):1108–1123.
4. Buch MH. Defining refractory rheumatoid arthritis. Ann Rheum Dis 2018;77(7):966–969.
5. Buch MH, Eyre S, McGonagle D. Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. Nat Rev Rheumatol 2021;17(1):17–33.
6. de Hair MJH, Jacobs JWG, Schoneveld JLM, et al. Difficult-to-treat rheumatoid arthritis: an area of unmet clinical need. Rheumatology (Oxford) 2018;57(7):1135–1144.
7. Matsson A, Solomon DH, Crabtree MM, et al. Patterns in the sequential treatment of rheumatoid arthritis patients starting a b/tsDMARD: 10-year experience from a US-based registry. Res Sq Preprint posted online February 27, 2023. doi:10.21203/rs.3.rs-2624931/v1
8. Nagy G, Roodenrijs NMT, Welsing PMJ, et al. EULAR points to consider for the management of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2022;81(1):20–33.
9. Roodenrijs NMT, de Hair MJH, van der Goes MC, et al; Whole EULAR Task Force on Development of EULAR Recommendations for the Comprehensive Management of Difficult-to-Treat Rheumatoid Arthritis. Characteristics of difficult-to-treat rheumatoid arthritis: results of an international survey. Ann Rheum Dis 2018;77(12):1705–1709.
10. Nagy G, Roodenrijs NMT, Welsing PM, et al. EULAR definition of difficult-to-treat rheumatoid arthritis. Ann Rheum Dis 2021;80(1):31–35.
11. Tan Y, Buch MH. 'Difficult to treat' rheumatoid arthritis: current position and considerations for next steps. RMD Open 2022;8(2):e002387.
12. Watanabe R, Hashimoto M, Murata K, et al. Prevalence and predictive factors of difficult-to-treat rheumatoid arthritis: the KURAMA cohort. Immunol Med 2022;45(1):35–44.
13. Watanabe R, Okano T, Gon T, et al. Difficult-to-treat rheumatoid arthritis: current concept and unsolved problems. Front Med (Lausanne) 2022;9:1049875.

14. Aletaha D, Neogi T, Silman AJ, et al. Rheumatoid arthritis classification criteria. Arthritis Rheum 2010;62(9):2569–2581.
15. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31(3):315–324.
16. Bykerk VP, Shadick N, Frits M, et al. Flares in rheumatoid arthritis: frequency and management. A report from the BRASS registry. J Rheumatol 2014;41(2):227–234.
17. Pappas DA, Kremer JM, Reed G, et al. Design characteristics of the CORRONA CERTAIN study: a comparative effectiveness study of biologic agents for rheumatoid arthritis patients. BMC Musculoskelet Disord 2014;15(1):113.
18. Paudel ML, Li R, Naik C, et al. Prevalence and characteristics of adults with difficult-to-treat rheumatoid arthritis in a large patient registry. Rheumatology (Oxford) 2025;64(3):1102–1110.
19. Prevoo ML, van 't Hof MA, Kuper HH, et al. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38(1):44–48.
20. Felson DT, Anderson JJ, Boers M, et al; The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36(6):729–740.
21. Pincus T, Callahan LF, Sale WG, et al. Severe functional declines, work disability, and increased mortality in seventy-five rheumatoid arthritis patients studied over nine years. Arthritis Rheum 1984;27(8):864–872.
22. Wolfe F, Michaud K, Pincus T, et al. The disease activity score is not suitable as the sole criterion for initiation and evaluation of anti-tumor necrosis factor therapy in the clinic: discordance between assessment measures and limitations in questionnaire use for regulatory purposes. Arthritis Rheum 2005;52(12):3873–3879.
23. Landewé R, van der Heijde D. Radiographic progression in rheumatoid arthritis. Clin Exp Rheumatol 2005;23(5 suppl 39):S63–S68.
24. van der Heijde D. How to read radiographs according to the Sharp/van der Heijde method. J Rheumatol 2000;27(1):261–263.
25. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann Rheum Dis 2001;60(9):817–827.
26. Lillegraven S, Prince FH, Shadick NA, et al. Remission and radiographic outcome in rheumatoid arthritis: application of the 2011 ACR/EULAR remission criteria in an observational cohort. Ann Rheum Dis 2012;71(5):681–686.
27. Pincus T, Yazici Y, Bergman MJ. RAPID3, an index to assess and monitor patients with rheumatoid arthritis, without formal joint counts: similar results to DAS28 and CDAI in clinical trials and clinical care. Rheum Dis Clin North Am 2009;35(4):773–778.
28. Pincus T, Summey JA, Soraci SA Jr, et al. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum 1983;26(11):1346–1353.
29. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024;385:e078378.
30. Iswaran H, Kogalur UB. Fast unified random forests for survival, regression, and classification (RF-SRC). 2025. https://cran.r-project.org/package=randomForestSRC
31. Ishwaran H, Kogalur UB, Blackstone EH, et al. Random survival forests. Ann Appl Stat 2008;2(3):841–860.
32. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Guyon I, von Luxburg U, Bengio S, et al, eds. Advances in Neural Information Processing Systems 30 (NIPS 2017). Curran Associates, Inc; 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
33. McWilliams DF, Dawson O, Young A, et al. Discrete trajectories of resolving and persistent pain in people with rheumatoid arthritis despite undergoing treatment for inflammation: results from three UK cohorts. J Pain 2019;20(6):716–727.
34. Bécède M, Alasti F, Gessl I, et al. Risk profiling for a refractory course of rheumatoid arthritis. Semin Arthritis Rheum 2019;49(2):211–217.
35. Messelink MA, Roodenrijs NMT, van Es B, et al. Identification and prediction of difficult-to-treat rheumatoid arthritis patients in structured and unstructured routine care data: results from a hackathon. Arthritis Res Ther 2021;23(1):184.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 66–75
DOI 10.1002/acr.25655

Trends in Opioid Prescriptions in Individuals With Axial Spondyloarthritis in the United States and United Kingdom in the Last Two Decades: Analysis of Electronic Health Records and Insurance Claims Data

H. Berk Degirmenci,[1] Christine E. Peloquin,[1] Maggie Westerland,[1] Sara Lodi,[2] Pedro M. Machado,[3,4] S. Reza Jafarzadeh,[1] Tuhina Neogi,[1,2] Lianne S. Gensler,[5] Maureen Dubreuil,[1] and Jean W. Liew[1]

Objective. Prolonged opioid use occurs in 25% of individuals with axial spondyloarthritis (axSpA). Whether introduction of tumor necrosis factor inhibitors (TNFi) influenced secular trends in opioid prescription in axSpA is unknown. We examined opioid prescription trends in axSpA, relative to nonsteroidal anti-inflammatory drugs (NSAIDs) and TNFi, using two observational databases.

Methods. We used data from IQVIA Medical Research Database (IMRD), a UK electronic health records–based database from 2000 to 2020, and Merative MarketScan (MarketScan), a US administrative claims-based database from 2006 to 2021. We included adults (18–89 years, IMRD; 18–65 years, MarketScan) with axSpA. We calculated annual prescription rates for opioids, NSAIDs, and TNFi (only in MarketScan) and estimated percentage changes in annual prescription rates by medication class using joinpoint regression.

Results. We included 1,689 individuals from IMRD (mean age 47 years; 74% male) and 18,858 individuals from MarketScan (mean age 45 years; 51% male). In IMRD, annual opioid prescription rates decreased by 2.5% (95% confidence interval [CI] –9.1 to 1.9) between 2001 and 2007, increased by 3.9% (95% CI –3.0 to 14.2) between 2007 and 2016, and decreased by 0.4% (95% CI –11.2 to 2.9) between 2016 and 2020. In MarketScan, annual opioid prescription rates decreased by 1.5% (95% CI –3.0 to 2.0) in 2008 to 2016 and decreased by 8.6% (95% CI –15.2 to –5.7) in 2016 to 2021. TNFi prescriptions increased by 1.8% (95% CI 1.1 to 2.5) in 2008 to 2021.

Conclusion. Opioid prescription rates remained stable over time in the United Kingdom, whereas they slightly decreased in the United States as TNFi uptake increased. These trends may reflect a US nationwide change in guidance for opioid prescriptions issued in 2016.

INTRODUCTION

Back pain is a predominant feature of axial spondyloarthritis (axSpA), which has a prevalence of 0.9% to 1.4% among adults in the United States and the United Kingdom.[1,2] AxSpA includes radiographic axSpA (r-axSpA; also known as ankylosing spondylitis) and nonradiographic axSpA (nr-axSpA), differentiated by the presence of sacroiliitis detectable on conventional radiography.[3] Despite the adverse effects of prolonged opioid use, including respiratory depression and the risk of opioid use disorder, up to quarter of people with axSpA in the United States are prescribed long-term opioids, typically defined as ≥90 cumulative days of use.[4–8]

The views expressed herein are those of the authors and do not necessarily represent those of the NHS, NIHR, or UK Department of Health.

Supported by the NIH (grant P30-AR-072571). Dr Machado's work was supported by the NIHR University College London Hospitals Biomedical Research Centre. Dr Jafarzadeh's work was supported by the Innovative Research Award from the Rheumatology Research Foundation. Dr Neogi's work was supported by the NIH (grant K24-AR-070892). Dr Dubreuil's work was supported by the Rheumatology Research Foundation and Boston University Clinical & Translational Science Institute. Dr Liew's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant K23-AR-082938); the Rheumatology Research Foundation Investigator Award; a Spondyloarthritis Research and Treatment Network pilot grant; and the Spondylitis Association of America Jane Bruckel Early Career Investigator Award.

[1]Chobanian & Avedisian School of Medicine, Boston University, Boston, Massachusetts; [2]School of Public Health, Boston University, Boston, Massachusetts; [3]University College London, London, United Kingdom; [4]NIHR University College London Hospitals Biomedical Research Centre, London, United Kingdom; [5]University of California, San Francisco.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25655).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25655.

Address correspondence via email to and Jean W. Liew, MD, MS, at jwliew@bu.edu.

Submitted for publication May 15, 2025; accepted in revised form September 10, 2025.

SIGNIFICANCE & INNOVATIONS

- Despite the introduction and availability of effective biologic therapies for axial spondyloarthritis (axSpA), including tumor necrosis factor inhibitors (TNFi), prolonged opioid use occurs in approximately 25% of this population. This ongoing reliance on opioids represents significant public health concern given their known adverse effects. Whether opioid use has declined in the two decades since the introduction and approval of TNFi for axSpA remains unclear from prior studies.
- To more comprehensively describe and evaluate annual prescription rates for opioids, nonsteroidal anti-inflammatory drugs (NSAIDs), and TNFi over time for individuals with axSpA, we used two large observational databases from the United Kingdom and the United States (IQVIA Medical Research Database and MarketScan, respectively) that included data spanning two decades. We employed joinpoint regression to detect changes over time in annual prescription rates and determine whether trends in opioid prescribing mirrored changes for NSAID and TNFi prescriptions.
- Our findings reveal that TNFi prescriptions increased over time in the United States, as expected. However, opioid prescription trends diverged, remaining stable in the United Kingdom but declining over time in the United States. This decline is more likely attributable to national opioid prescribing guidelines introduced by the US Centers for Disease Control and Prevention in 2016 rather than a direct consequence of higher TNFi uptake among individuals with axSpA.

Long-term opioid use prevalence is higher in axSpA compared to other rheumatic diseases such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), which have estimated prevalences of long-term opioid use of 19% and 16%, respectively.[9] Contributing factors to long-term opioid use in axSpA may include a diagnostic delay averaging around six years, persistent disease activity despite appropriate treatment, and lower likelihood of initial tumor necrosis factor inhibitor (TNFi) continuation among opioid users.[10,11] Although treatment options for axSpA have improved over the past two decades, including the introduction and approval of TNFi in the 2000s, opioid use remains common. A US population-based study identified back pain as a primary reason for opioid use among individuals with musculoskeletal conditions.[12] Although the US Centers for Disease Control and Prevention (CDC) recommend limiting prolonged opioid use for noncancer pain, specific recommendations regarding opioid prescription in axSpA are lacking in guidelines from the American College of Rheumatology (ACR), EULAR, and Pan American League of Associations for Rheumatology.[3,4,7,13,14]

Despite the increased prescription of biologics for axSpA since their approval, a US claims-based study[15] showed that opioid prescriptions did not decline between 2006 and 2016. A recent UK electronic medical record (EMR)–based study also reported stable trends in opioid prescriptions among patients with axSpA.[16] These findings are contrary to the expectation that, with rising prescriptions and more common usage of TNFi for axSpA, opioid prescriptions should have declined with improved control of disease activity and thus of chronic back pain. Although existing studies in the United Kingdom and the United States assessing opioid use in axSpA evaluated changes in opioid prescriptions, neither were able to fully address the extent to which opioid prescription rates may have changed since TNFi approval for axSpA in both countries and whether there was a difference in trends between the two countries.

To address the knowledge gap from prior studies, we aimed to investigate trends in annual prescription rates for opioids among individuals with axSpA over a specified time period that included the introduction of biologic therapies for axSpA, specifically TNFi. We leveraged two different observational data sources spanning two decades: IQVIA Medical Research Database (IMRD), an EMR-based database from the United Kingdom, and Merative MarketScan (MarketScan), a claim-based database from the United States. In contrast to prior studies indicating stability in opioid prescriptions, we hypothesized that opioid prescriptions would decrease following the introduction and rising uptake of TNFi therapies in both the United Kingdom and the United States.

METHODS

Study design and data sources. We conducted an ecological study using two different observational databases from the United Kingdom and the United States. IMRD, incorporating THIN, a Cegedim database, is an EMR-based database representative of the UK population. IMRD includes data from more than 17 million patient records. The database holds patient demographic information, prescribed medication, diagnoses, laboratory test results, and additional information such as lifestyle factors, body mass index (BMI), and vaccinations as recorded in general practices.[17] The use of IMRD for research was approved by the NHS Health Research Authority (NHS Research Ethics Committee ref 23/EM/0151) for medical and public health research.[18] This study received Scientific Review Committee approval (21SRC053).

The Merative MarketScan Databases (referred to as MarketScan hereafter) is one of the largest collections of deidentified patient-level claims data in the United States on approximately 273 million unique patients with employer-sponsored insurance and their dependents. The database includes encounter-level services data from outpatient, emergency department, inpatient, pharmacy, and laboratory claims.[19]

Study population. *Inclusion.* In IMRD, we identified a cohort of adults with axSpA aged 18 and 89 years enrolled in a general practitioner (GP) practice from 2000 to 2020. In MarketScan, we identified a cohort of adults with axSpA aged 18 to 65 years from 2006 to 2021. AxSpA was defined as at least two diagnostic Read codes for ankylosing spondylitis seven or more days apart in IMRD[20] and as at least two *International Classification of Diseases, Ninth or Tenth Revision* (ICD-9 or ICD-10) diagnostic codes for ankylosing spondylitis seven or more days apart in MarketScan. This definition of axSpA was previously found to have an 89% positive predictive value (indicating their ability to correctly identify a condition when the condition is present) for identifying ankylosing spondylitis (r-axSpA).[21] For this study, we included only these diagnostic codes because over the course of our study period, the available diagnostic codes were for ankylosing spondylitis (r-axSpA), and individuals with nr-axSpA would have been labeled with the same codes by clinicians. The specific ICD-10 code for nr-axSpA (M45.A) was officially approved[21–23] on October 1, 2021. Finally, we required continuous enrollment for two or more years before patients met inclusion criteria for axSpA to allow for the measurement of demographic and clinical variables.

Exclusion. To mitigate potential misclassification of axSpA, we excluded individuals with RA, identified by at least two diagnostic (Read or ICD) codes within two years before axSpA diagnosis.[24] To address indications for prolonged opioid use other than axSpA, including the use of opioids for pain management in individuals with a history of malignancy, we excluded individuals with any diagnostic codes for malignancy (excluding benign tumors and nonmelanoma skin cancer) within two years before the axSpA diagnosis date, defined as the second qualifying diagnostic code for axSpA. This also reflects exclusion criteria from a prior study by Jani et al, which examined opioid prescription trends using the UK Clinical Practice Research Datalink (CPRD).[25] We were not able to specifically identify indications for opioid use such as surgery or trauma for exclusion (see Supplementary Items A and B for Consolidated Standards of Reporting Trials (CONSORT) diagrams depicting the data selection process).

Outcomes. In both databases, we extracted prescription- and person-level information, including the drug name, formulation, route of administration, dosage, quantity, and prescription dates (IMRD) or dispense dates (MarketScan). Opioids were our primary medication class of interest; for the comparators, we extracted data related to nonsteroidal anti-inflammatory drugs (NSAIDs) (IMRD and MarketScan) and TNFi (available in MarketScan only). We did not include conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) as a comparator because they are not recommended in the management of axial disease in spondyloarthritis.[3,13,14] Follow-up began at the axSpA diagnosis date and ended at the earliest date of a code for malignancy, end of life, palliative care, hospice care (per previously published algorithms[26]), death, or database closure (December 2020 for IMRD; December 2021 for MarketScan). For IMRD, follow-up could also end at age 90 years, with the individual transferring out of their GP practice, or on the last date the GP contributed data to IMRD.

Medications of interest. *Opioids.* We included oral formulations for opioids, including tramadol. We excluded methadone because it is primarily used for the treatment of substance use disorders.[25] We included topical formulations of fentanyl and excluded intravenous and intranasal formulations (see full list in Supplementary Item C).

NSAIDs and TNFi. We included oral formulations of selective cyclooxygenase 2 inhibitors and nonselective NSAIDs (see full list in Supplementary Item C). Biologic prescriptions are poorly captured in IMRD because this database does not capture consultant (specialist) data, and GPs do not routinely prescribe biologics. Thus, we extracted information about TNFi prescriptions from MarketScan only. TNFi included adalimumab, certolizumab, etanercept, golimumab, and infliximab. TNFi use was determined through a combination of pharmacy claims and *Current Procedural Terminology* codes on medical claims for infusions.

Annual prescription rates. In IMRD and MarketScan separately, we calculated annual prescription rates (the number of opioid prescriptions divided by the total person-years for each calendar year) for opioids and for NSAIDs (IMRD and MarketScan) and TNFi (available in MarketScan only) as comparators.

Opioid strength (daily morphine milligram equivalents). The strength of prescribed opioids was quantified by daily morphine milligram equivalents (MMEs). For each opioid prescription, the daily dose was multiplied by the MME conversion factor for the corresponding drug group.[4,27,28] When the daily dose information was missing, the imputed median daily dose for that opioid type was used. For example, an oxycodone 5-mg prescription (correlating to an MME of 7.5) with a daily dose of four times daily was calculated as follows: a daily dose of four multiplied by an MME conversion factor of 7.5, resulting in a daily MME of 30. We reported annual trends for opioid daily MMEs using prescription-level data in IMRD and MarketScan separately (see Supplementary Items D and E for further information).

Opioid duration (days' supply). The duration of opioid prescriptions was assessed as days' supply per prescription. For each opioid prescription, days' supply data were directly extracted from MarketScan. In IMRD, days' supply was calculated from prescription quantity and daily dose. On prescriptions with missing daily dose, the daily dose was imputed to the mean daily dose for the opioid drug type. We reported annual trends for opioid prescription median days' supply using prescription-

level data in IMRD and MarketScan separately (see Supplementary Items D and F for further information).

Covariates. To describe the study population, we extracted demographic and clinical characteristics, including age, sex, other treatments for axSpA, and the presence of comorbidities relevant to both opioid use and axSpA. These comorbidities included coronary artery disease (CAD), chronic kidney disease (CKD), diabetes mellitus, hypertension, tobacco use, alcohol use, and obesity (defined as BMI ≥30). Comorbidities were defined by the presence of at least one diagnostic code any time before the axSpA diagnosis date. Treatments for axSpA (eg, glucocorticoids, csDMARDs, TNFi) were defined as any prescription within one year before the axSpA diagnosis date. We also extracted prior opioid use, defined as any prior opioid prescription within two years before the axSpA diagnosis date.

Statistical analysis. Based on the data available in IMRD, we examined outcomes between 2001 and 2020. In MarketScan, we examined outcomes between 2008 and 2021.

Joinpoint regression analysis. To examine the trends in annual opioid prescription rates, we used joinpoint regression analysis, using the method recommended by the National Cancer Institute and the Joinpoint Trend Analysis software (v.5.2.0).[29] Joinpoint regression is a data-driven statistical method used to identify points (joinpoints) where a trend changes significantly. It allows us to estimate the potential number and locations of joinpoints based on the Bayesian information criterion, a penalized likelihood criterion that balances model fit and complexity. It involves Monte Carlo permutations to assess the statistical significance of the identified joinpoints by comparing the observed fit to a distribution of fits obtained from random permutations of the data.[30–32] Under a Poisson distribution, the model allows us to estimate the annual percent change (APC) for each segment that quantified the rate of change in prescription rates (opioids, NSAIDs, and TNFi separately). The average APC (AAPC), which represents the overall trend over a prespecified time interval (eg, the entire study period) and is valid without assuming linearity in trends over time (ie, even when the trend changes joinpoints), was also calculated for the full time period for prescription rates to provide a single summary estimate of the overall trend and to complement the segment-specific APCs.

Sensitivity analyses. Although the joinpoint regression models agnostically estimate the optimal number of joinpoints for each medication's annual prescription rate trend, results may not be clinically relevant or easily interpretable. As a sensitivity analysis, we prespecified that models include one joinpoint for each medication category for joinpoint regression analyses in IMRD and MarketScan, with the assumption that two time trends may be more clinically interpretable.

Ethical approval. Because this study involved the analysis of pre-existing deidentified data, it was deemed exempt from requiring institutional review board approval (Boston University Medical Campus IRB H-42058).

RESULTS

Demographic and clinical characteristics. In the UK IMRD cohort, we identified 1,689 eligible individuals with axSpA (mean age 47 years, 74% male, median follow-up time of 5.6 years [interquartile range (IQR) 2.5–9.9 years]). Of these, 18.9% had hypertension, 4.6% had diabetes mellitus, 5.9% had CAD, and 3.4% had CKD. csDMARD use was present in 10.5%, glucocorticoid use was present in 10.5%, NSAID use was present in 71.7%, and any prior opioid use (in the prior two years) was present in 54.1% (Table 1). In the US MarketScan cohort, we identified 18,858 eligible individuals with axSpA (mean age 45 years, 51% male, median follow-up time of 1.7 years [IQR 0.7–3.3 years]). Comorbidities were more common compared to the UK cohort: 38.8% had hypertension, 14.2% had diabetes mellitus, 13.3% had CAD, and 8.6% had CKD. csDMARD use was present in 13.6%, and any prior opioid use was present in 53.8%, which were similar to the UK cohort, though there was more glucocorticoid use (40.3%) and less NSAID use (52.3%) (Table 1).

Annual prescription rates. In the IMRD cohort, the annual opioid prescription rate per person-year was 3.24 in 2001 and 3.86 in 2020 (Supplementary Item G). Joinpoint

Table 1. Demographic and clinical characteristics from IMRD (United Kingdom) and MarketScan (United States) databases*

	IMRD (2000–2020)	MarketScan (2006–2021)
Number of individuals	1,689	18,858
Mean age (SD), y	47 (14.1)	45 (12)
Male, %	74	51
Median follow-up time (IQR), y	5.6 (2.5–9.9)	1.7 (0.7–3.3)
Hypertension, %	18.9	38.8
Diabetes mellitus, %	4.6	14.2
Coronary artery disease, %	5.9	13.3
Chronic kidney disease, %	3.4	8.6
Current/former tobacco use, %	48.5	10.9
Current/former alcohol use, %	68.0	3.0
Obesity, %	17.6	19.6
csDMARD use, %	10.5	13.6
Glucocorticoid use, %	10.5	40.3
NSAID use, %	71.7	52
Any prior opioid use, %	54.1	53.8%

* Obesity was defined as body mass index of 30 or higher. Comorbidities were captured any time before the axSpA diagnosis date. Medications were captured within one year before the axSpA diagnosis date. Any prior opioid use was captured within two years before the axSpA diagnosis date. axSpA, axial spondyloarthritis; csDMARD, conventional synthetic disease-modifying antirheumatic drug; IMRD, IQVIA Medical Research Database; IQR, interquartile range; NSAID, nonsteroidal anti-inflammatory drug.

regression analysis identified two inflection points resulting in three segments for opioids, showing a decreasing trend by 2.5% annually (APC –2.5, 95% confidence interval [CI] –9.1 to 1.9) from 2001 to 2007, a rising trend by 3.9% annually (APC 3.9, 95% CI –3.0 to 14.2) from 2007 to 2016, and lastly a near-stable downtrend in opioid prescription rates by 0.4% (APC –0.4, 95% CI –11.2 to 2.9) from 2016 to 2020 (Figure 1, Table 2), with non–statistically significant changes in all three segments. The annual NSAID prescription rate per person-year was 5.72 in 2001 and 2.68 in 2020. Joinpoint regression analyses identified three joinpoints resulting in four segments for NSAIDs, showing a statistically significant decrease by 6.5% annually (APC –6.5, 95% CI –10.5 to –2.3) between 2001 and 2006, then a significant decrease by 3.7% annually (APC –3.7, 95% CI –8.1 to –0.4) between 2006 and 2010, a nonsignificant decrease by 1.1% annually (APC –1.1, 95% CI –7.0 to 2.3) between 2010 and 2015, and finally a significant decrease by 4.1% (APC –4.1, 95% CI –7.9 to –0.1) between 2015 and 2020. For the entire time period of 2001 to 2020, the AAPC for opioids in IMRD reflected a statistically significant increase of 0.9% annually (AAPC 0.9, 95% CI 0.8 to 1.6), whereas the AAPC for NSAIDs showed a significant decrease of 3.9% annually (AAPC –3.9, 95% CI –4.2 to –3.6).

In the MarketScan cohort, the annual opioid prescription rate per person-year decreased by 2.97 in 2008 to 1.74 in 2021 (Supplementary Item H). Joinpoint regression identified one inflection point resulting in two segments for opioids, showing a non–statistically significant decrease of 1.5% annually (APC –1.5, 95% CI –3.0 to 2.0) from 2008 to 2016, followed by a significant decline of 8.6% annually (APC –8.6, 95% CI –15.2 to –5.7) from 2016 to 2021 (Figure 2, Table 2). NSAID prescriptions also statistically significantly declined from 1.92 in 2008 to 1.60 in 2021, with a 1.1% annual decrease (APC –1.1, 95% CI –1.9 to –0.2) over the entire period, without identified joinpoints. TNFi prescriptions statistically significantly increased from 1.91 in 2008 to 2.46 in 2021, showing a 1.8% annual increase (APC 1.8, 95% CI 1.1 to 2). For the entire time period of 2008 to 2021, the AAPC for opioids in MarketScan showed a statistically significant decrease of 4.3% annually (AAPC –4.3, 95% CI –5.5 to –3.1), whereas NSAIDs significantly declined by 1.1% annually (AAPC –1.1, 95% CI –1.9 to –0.2), and TNFi significantly increased by 1.8% annually (AAPC 1.8, 95% CI 1.2 to 2.5). Results were similar in sensitivity analyses for both cohorts using prespecified models for one joinpoint for each medication category (Supplementary Items I and J).

Opioid duration and strength. For both cohorts, opioid prescription days' supply and daily MMEs were numerically higher at the end of the study period compared to the start. In IMRD, the median days' supply for opioid prescriptions was 17 days in 2001 and 22 days in 2020. In MarketScan, the median days' supply for opioid prescriptions was 13 days in 2008 and 30 days in 2021. In IMRD, the median opioid daily MME was

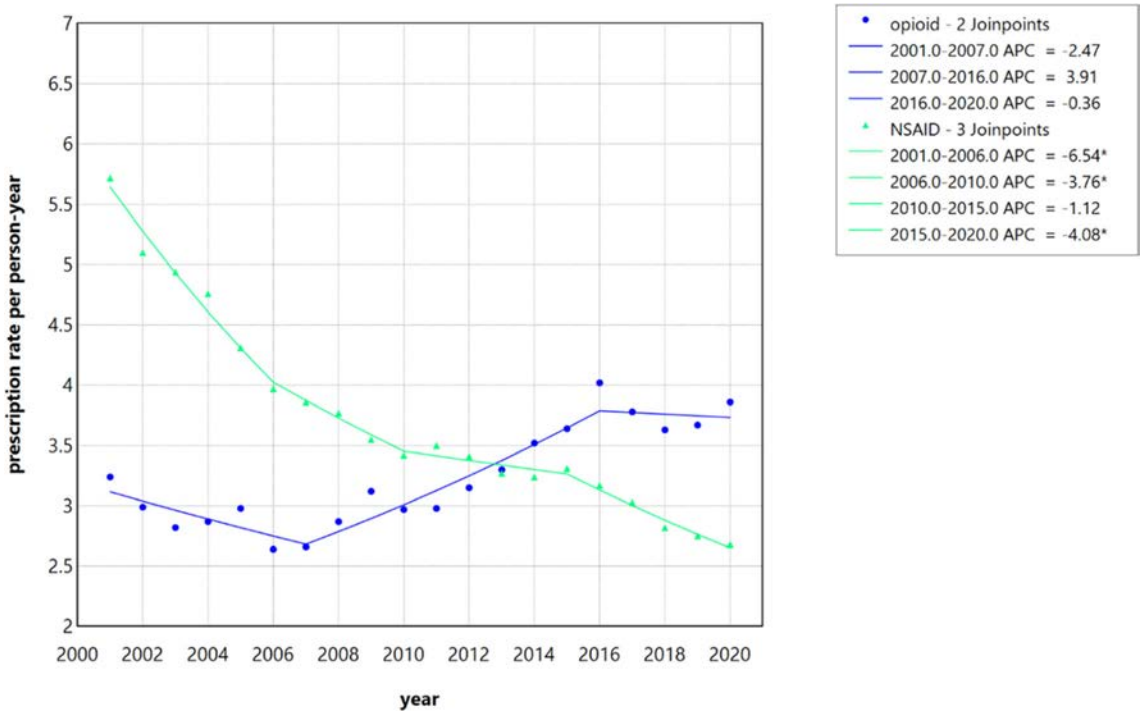


Figure 1. Opioid annual prescription rates (per person-year) over time in IMRD (United Kingdom). The figure shows the plot of the data points for opioid and NSAID prescription rates by year, with regression lines fit by joinpoint regression. The slope of each segment's regression line is the APC. Joinpoints are identified by model fit statistics for when trends change. APC, annual percent change for prescription rate per person-year; IMRD, IQVIA Medical Research Database; NSAID, nonsteroidal anti-inflammatory drug. *, statistically significant.

Table 2. Trends in medication prescription rates by data source*

Dataset (y)	Prescription rate (start)	Prescription rate (end)	AAPC% (95% CI)	Trend 1, APC% (95% CI)	Trend 2, APC% (95% CI)	Trend 3, APC% (95% CI)	Trend 4, APC% (95% CI)
Opioids							
IMRD (2001–2020)	3.24	3.86	0.9 (0.8 to 1.6)	2001–2007: −2.5 (−9.1 to 1.9)	2007–2016: 3.9 (−3.0 to 14.2)	2016–2020: −0.4 (−11.2 to 2.9)	n/a
MarketScan (2008–2021)	2.97	1.74	−4.3 (−5.5 to −3.1)	2008–2016: −1.5 (−3.0 to 2.0)	2016–2021: −8.6 (−15.2 to −5.7)	n/a	n/a
NSAIDs							
IMRD (2001–2020)	5.72	2.68	−3.9 (−4.2 to −3.6)	2001–2006: −6.5 (−10.5 to −2.3)	2006–2010: −3.7 (−8.1 to −0.4)	2010–2015: −1.1 (−7.0 to 2.3)	2015–2020: −4.1 (−7.9 to −0.1)
MarketScan (2008–2021)	1.92	1.60	−1.1 (−1.9 to −0.3)	2008–2021: −1.1 (−1.9 to −0.2)	n/a	n/a	n/a
TNFi							
MarketScan (2008–2021)	1.91	2.46	1.8 (1.2 to 2.5)	2008–2021: 1.8 (1.1 to 2.5)	n/a	n/a	n/a

* The prescription rate was defined as the number of prescriptions per person-year. Person-years were calculated for each individual by summing the total duration of follow-up, from the date of axSpA diagnosis to the end of the study, the last available follow-up date, or the use of any diagnostic codes for the following: malignancy, end of life, palliative care, or hospice care. Results of joinpoint regression analyses for each medication class and each data source are presented as the APC with 95% CI (rate of change in prescription rates) for each segment (trend). The AAPC represents the trend over the entire time period as a single summary estimate of the overall trend. AAPC, average annual percent change; APC, annual percent change for prescription rate per person-year; axSpA, axial spondyloarthritis; CI, confidence interval; IMRD, IQVIA Medical Research Database; n/a, not applicable; NSAID, nonsteroidal anti-inflammatory drugs; TNFi, tumor necrosis factor inhibitors.

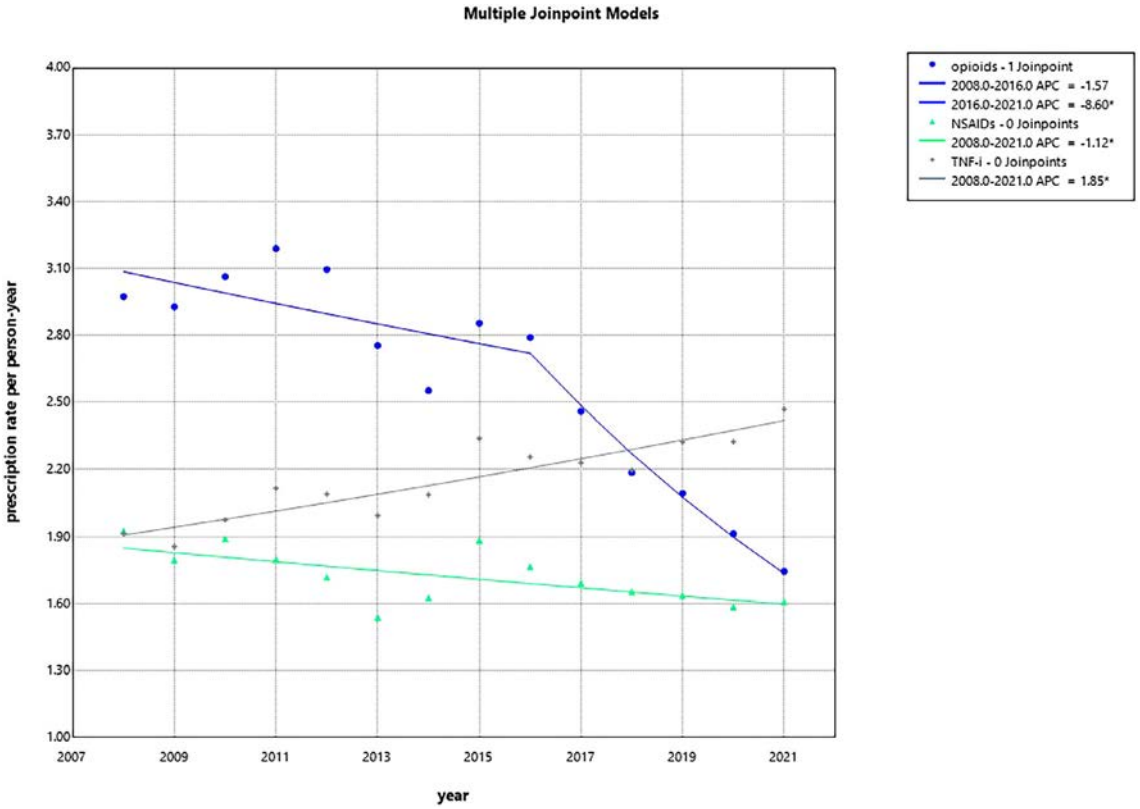


Figure 2. Opioid annual prescription rates (per person-year) over time in MarketScan (United States). The figure shows the plot of the data points for opioid, NSAID, and TNFi prescription rates by year, with regression lines fit by joinpoint regression. The slope of each segment's regression line is the APC. Joinpoints are identified by model fit statistics for when trends change. APC, annual percent change for prescription rate per person-year; NSAID, nonsteroidal anti-inflammatory drug; TNF-i, tumor necrosis factor inhibitor. *, statistically significant. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25655/abstract.

20 MME/day in 2001 and 27 MME/day every year from 2002 to 2020. In MarketScan, the median opioid daily MME was 7 MME/day in 2008 and 22.5 MME/day in 2021 (Supplementary Items E, F, K, and L).

DISCUSSION

In this analysis of two large observational databases from the United Kingdom and the United States, we observed that opioid prescription rates for patients with axSpA were about three per person-year and remained stable in the United Kingdom, whereas they were slightly declining in the United States, particularly after 2016. This decrease in the United States might likely be attributed to the 2016 CDC guidelines, which prompted a broad reduction in opioid prescriptions nationwide.[7,33] Furthermore, the fall in opioid prescriptions in both countries was not directly paralleled with a commensurate increase in TNFi prescriptions in the United States, again supporting the notion that these represent secular trends of prescribing practices rather than decreases in chronic pain prevalence.

AxSpA, which affects approximately 1% of US adults, shares a similar prevalence with that of RA.[1] Opioid use is particularly more common in axSpA compared to other rheumatic diseases, and it is associated with increased health care use, including, but not limited to, more hospitalizations, acute care, and outpatient visits.[9,34] Chronic pain and functional impairment in axSpA contribute to a high frequency of opioid use, about 25% according to US data[5] from 2012 to 2017. Opioid use in this group leads to higher health care use and doubling of medical costs.[34,35] When compared to age- and sex-matched patients with hypertension, patients with axSpA have a higher relative risk of long-term opioid use compared to those with other rheumatic diseases such as RA and SLE (relative risk of 2.73 in axSpA vs 2.21 [RA] and 1.82 [SLE]).[9]

Although opioids provide symptomatic relief, they do not address underlying inflammation, potentially delaying diagnosis and contributing to higher rates of TNFi treatment failure.[5,6,11] Opioid users with axSpA are less likely to persist on TNFi treatment, as seen in both MarketScan and Optum Clinformatics studies.[11,36] The 2019 ACR/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network guideline and the 2022 Assessment of SpondyloArthritis international Society (ASAS)-EULAR recommendations caution against long-term opioid use but do not offer specific recommendations.[3,13] A systematic review highlighted the lack of evidence supporting strong

opioids in managing inflammatory arthritis pain, advising caution due to the risks associated with prolonged use.[37]

In our study, we observed a steady rise in TNFi prescriptions in the United States from 2008 to 2021, increasing by nearly 2% per year over this period and reflecting the growing use of TNFi in axSpA management. With the increasing use of biologics for axSpA, one might expect concurrent improvements in chronic pain management. A Finnish registry-based study did report reduced opioid use following biologic DMARD initiation,[12] though differences in health care systems may limit generalizability of these findings to UK and US populations. In the United Kingdom, reductions in NSAID use have outpaced declines in opioid prescribing[16]—an effect that may be partly attributed to the 2004 guidance issued by the UK Medicines and Healthcare Products Regulatory Agency on NSAID safety,[38] consistent with our results. Our study found that opioid prescribing in the United Kingdom since 2016 has remained relatively stable for individuals with axSpA, mirroring trends in the general population.[39] A US MarketScan study using data from 2006 to 2016 also showed stable opioid prescriptions for patients with axSpA, whereas TNFi use increased and NSAID use decreased.[7,15] Our study, which includes more recent data since 2016, showed a slight subsequent decline in US opioid prescriptions, likely reflecting the impact of the CDC's 2016 guidelines aimed at reducing opioid use nationally.[33]

Despite this decline in prescription rates over time, we noted longer opioid prescription durations and increased strength as measured by daily MME. The 2016 CDC clinical practice guideline[7] for prescribing opioids and the 2022 ASAS-EULAR recommendations[13] for axSpA caution against long-term opioid use for chronic pain. The 2022 CDC clinical practice guideline[4] for prescribing opioids highlights that the risk of long-term opioid use at follow-up after initiation for short-term use for acute pain might be greater with higher dosages and a longer initial duration of exposure, calling for careful interpretation of the slight decline in opioid prescriptions in the United States after 2016. This modest decline in prescription rates may be offset by prescriptions of longer duration and greater strength. Both UK and US cohorts showed a numerical increase in opioid prescription strength over time, with our study demonstrating an average daily MME lower than the UK CPRD study by Huang et al,[8] which reported daily MMEs of 34 to 35. Despite minor differences, these studies highlight the increasing strength of opioids prescribed to patients with axSpA, raising concerns about the potential for opioid-related adverse events. Prolonged opioid use at higher doses is associated with greater risks of emergency visits, hospital readmissions, and death.[40,41] Additionally, the general UK population saw a significant rise in oral morphine equivalency during a similar period.[39]

The reasons behind persistent opioid use in axSpA remain unclear. The Prospective Study of Outcomes in Ankylosing Spondylitis (PSOAS) identified older age, comorbidities, functional impairment, and subjective disease activity as risk factors for prolonged opioid use, with an inverse (but not statistically significant) association with TNFi usage.[6] A more recent UK CPRD study found similar risk factors, indicating that noninflammatory pain driven by comorbidities or chronic damage from inflammatory arthritis may contribute to opioid prescriptions.[12] Despite improvements in treatment, significant unmet needs persist in managing chronic pain in axSpA, including, but not limited to, addressing fibromyalgia in patients with axSpA, which is estimated to be around[42] 20%.

Our study has several limitations. First, it is observational, using administrative claims (MarketScan, United States) and GP-entered records (IMRD, United Kingdom), each with specific constraints. MarketScan excludes individuals without commercial insurance, limiting generalizability to the full US population, such as those on Medicare or Medicaid, whereas IMRD lacks information on biologic prescriptions. Both data sources lack information on axSpA disease duration, severity, or activity, as well as functional measures. We used similar algorithms to identify axSpA cohorts in both databases, but coding may vary between data sources. Second, misclassification of axSpA based on diagnostic codes is possible, potentially leading to underestimation and/or overestimation of disease prevalence and baseline characteristics, although we employed a validated approach to minimize this.[21] By excluding individuals with RA codes, specificity for axSpA was increased at the cost of sensitivity. Third, opioid indications and prescribing specialties were not captured, but we excluded individuals with conditions that typically require opioids, such as malignancy. We did not include methadone prescriptions given its use in substance use disorder. However, we could not identify non–axSpA-related indications such as surgery and trauma for opioid prescriptions for exclusion. Although we considered overcoming this potential bias by analyzing short-acting versus long-acting formulations separately, this stratification was insufficient to fully distinguish short-term from long-term indications for opioid use across databases. Fourth, over-the-counter NSAID use cannot be accurately assessed in either the US or the UK database, leading to potential underestimation of NSAID use. Fifth, we did not directly compare opioid use with that of the general populations in the United States and United Kingdom but have referenced prior literature for indirect comparisons. Finally, this descriptive study was intended to examine opioid use trends at the ecological level; therefore, individual-level adjustments were not performed, which precludes our ability to assess the influence of concomitant conditions or the impact of changes in patient characteristics over time on prescription rate trends.

The strength of our study lies in its assessment of longitudinal trends in opioid, NSAID, and TNFi prescription rates across UK and US data. Our study period spans the increasing use of biologic therapies for axSpA, allowing for comparison of opioid and NSAID trends. Joinpoint regression analyses identified inflection points in prescription trends, and we additionally reported trends

in opioid prescription durations and strengths, providing a comprehensive assessment of opioid use burden.

Our findings highlight that despite increased TNFi use over the past two decades, opioid prescribing trends have remained stable in the United Kingdom and have only modestly declined in the United States, likely due to national opioid prescribing policies rather than biologic uptake. These results underscore the need for continued efforts to better understand and address persistent opioid use in axSpA, even in the context of improved disease-modifying treatment options. Continuing to identify and address modifiable factors driving prolonged opioid use in this population is crucial.

ACKNOWLEDGMENTS

The authors acknowledge the Boston Musculoskeletal Clinical Research Collaboratory Research Accelerator group for their insightful comments on this article. The authors also thank Kiryn Denmead for her assistance with optimization of the figures presented in the main article and Supplementary Material.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Reveille JD, Witter JP, Weisman MH. Prevalence of axial spondylarthritis in the United States: estimates from a cross-sectional survey. Arthritis Care Res (Hoboken) 2012;64(6):905–910.
2. Hamilton L, Macgregor A, Toms A, et al. The prevalence of axial spondyloarthritis in the UK: a cross-sectional cohort study. BMC Musculoskelet Disord 2015;16(1):392.
3. Ramiro S, Nikiphorou E, Sepriano A, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. Ann Rheum Dis 2023;82(1):19–34.
4. Dowell D, Ragan KR, Jones CM, et al. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022;71(3):1–95.
5. Sloan VS, Sheahan A, Stark JL, et al. Opioid use in patients with ankylosing spondylitis is common in the United States: outcomes of a retrospective cohort study. J Rheumatol 2019;46(11):1450–1457.
6. Dau JD, Lee M, Ward MM, et al. Opioid analgesic use in patients with ankylosing spondylitis: an analysis of the prospective study of outcomes in an ankylosing spondylitis cohort. J Rheumatol 2018;45(2):188–194.
7. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain--United States, 2016. JAMA 2016;315(15):1624–1645.
8. Huang YT, Jenkins DA, Yimer BB, Jani M. Factors associated with long-term opioid use among patients with axial spondyloarthritis or psoriatic arthritis who initiate opioids. Rheumatology (Oxford) 2025;64(4):1844–1852.
9. Chen SK, Feldman CH, Brill G, et al. Use of prescription opioids among patients with rheumatic diseases compared to patients with hypertension in the USA: a retrospective cohort study. BMJ Open 2019;9(6):e027495.
10. Redeker I, Callhoff J, Hoffmann F, et al. Determinants of diagnostic delay in axial spondyloarthritis: an analysis based on linked claims and patient-reported survey data. Rheumatology (Oxford) 2019;58(9):1634–1638.
11. Hunter T, Schroeder K, Sandoval D, et al. Persistence, discontinuation, and switching patterns of newly initiated TNF inhibitor therapy in ankylosing spondylitis patients in the United States. Rheumatol Ther 2019;6(2):207–215.
12. Muilu P, Rantalaiho V, Kautiainen H, et al. Opioid use frequency in early axial spondyloarthritis in Finland - a pharmacoepidemic register study. Joint Bone Spine 2022;89(3):105302.
13. Ward MM, Deodhar A, Gensler LS, et al. 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. Arthritis Care Res (Hoboken) 2019;71(10):1285–1299.
14. Bautista-Molano W, Fernández-Ávila DG, Brance ML, et al. Pan American League of Associations for Rheumatology recommendations for the management of axial spondyloarthritis. Nat Rev Rheumatol 2023;19(11):724–737.
15. Walsh J, Hunter T, Schroeder K, et al. Trends in diagnostic prevalence and treatment patterns of male and female ankylosing spondylitis patients in the United States, 2006–2016. BMC Rheumatol 2019;3(1):39.
16. Scott IC, Bailey J, White CR, et al. Analgesic prescribing in patients with inflammatory arthritis in England: an observational study using electronic healthcare record data. Rheumatology (Oxford) 2022;61(8):3201–3211.
17. IQVIA Medical Research Data (IMRD). Accessed August 1, 2024. https://www.iqvia.com/locations/united-kingdom/solutions/life-sciences-industry-solutions/real-world-solutions/iqvia-medical-research-data
18. IQVIA Medical Research Data. Accessed August 1, 2024. https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/iqvia-medical-research-data/
19. Merative MarketScan Research Databases. Accessed August 1, 2024. https://www.merative.com/documents/brief/marketscan-explainer-general
20. Benson T. The history of the Read Codes: the inaugural James Read Memorial Lecture 2011. Inform Prim Care 2011;19(3):173–182.
21. Dubreuil M, Peloquin C, Zhang Y, et al. Validity of ankylosing spondylitis diagnoses in The Health Improvement Network. Pharmacoepidemiol Drug Saf 2016;25(4):399–404.
22. Beresford L. Non-radiographic axial spondyloarthritis recognized with ICD-10 code. August 21, 2020. Accessed August 28, 2024. https://www.the-rheumatologist.org/article/non-radiographic-axial-spondyloarthritis-recognized-with-icd-10-code/
23. Department of Health & Human Services (DHHS), Centers for Medicare & Medicaid Services (CMS). Pub 100-04 Medicare claims processing. Accessed September 15, 2024. https://www.cms.gov/files/document/r10996cp.pdf
24. Curtis JR, Xie F, Smith C, et al. Changing trends in opioid use among patients with rheumatoid arthritis in the United States. Arthritis Rheumatol 2017;69(9):1733–1740.
25. Jani M, Birlie Yimer B, Sheppard T, et al. Time trends and prescribing patterns of opioid drugs in UK primary care patients with non-cancer

pain: a retrospective cohort study. PLoS Med 2020;17(10): e1003270.

26. Hua M, Li G, Clancy C, et al. Validation of the V66.7 code for palliative care consultation in a single academic medical center. J Palliat Med 2017;20(4):372–377.
27. Opioid oral morphine miligram equivalent (MME) conversion factors. Accessed October 5, 2025. https://medicaid-documents.dhhs.utah.gov/Documents/files/Opioid-Morphine-EQ-Conversion-Factors.pdf
28. Opioid equivalence chart. Accessed August 1, 2024. gloshospitals.nhs.uk
29. *Joinpoint Regression Program*. Version 5.2.0. Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute; June 2024. Accessed August 25, 2024. https://surveillance.cancer.gov/joinpoint/
30. Hincapie-Castillo JM, Goodin A. Using joinpoint regression for drug utilization research: tutorial and case study of prescription opioid use in the United States. Pharmacoepidemiol Drug Saf 2023;32(5): 509–516.
31. Chen HS, Zeichner S, Anderson RN, et al. The joinpoint-jump and joinpoint-comparability ratio model for trend analysis with applications to coding changes in health statistics. J Off Stat 2020;36(1):49–62.
32. Kim HJ, Fay MP, Feuer EJ, et al. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med 2000;19(3): 335–351.
33. Bohnert ASB, Guy GP Jr, Losby JL. Opioid prescribing in the United States before and after the Centers for Disease Control and Prevention's 2016 opioid guideline. Ann Intern Med 2018;169(6): 367–375.
34. Liew JW, Pedro S, Michaud K, et al. Opioid use and healthcare utilization in adults with psoriatic arthritis and ankylosing spondylitis. J Rheumatol 2024;51(10):1009–1014.
35. Boonen A, Brinkhuizen T, Landewé R, et al. Impact of ankylosing spondylitis on sick leave, presenteeism and unpaid productivity, and estimation of the societal cost. Ann Rheum Dis 2010;69(6):1123–1128.
36. Katz G, Ogdie A, Baker JF, et al. Association between depression, anxiety, chronic pain, or opioid use and tumor necrosis factor inhibitor persistence in inflammatory arthritis. Clin Rheumatol 2022;41(5): 1323–1331.
37. Whittle SL, Colebatch AN, Buchbinder R, et al. Multinational evidence-based recommendations for pain management by pharmacotherapy in inflammatory arthritis: integrating systematic literature research and expert opinion of a broad panel of rheumatologists in the 3e Initiative. Rheumatology (Oxford) 2012;51(8):1416–1425.
38. Bedson J, Belcher J, Martino OI, et al. The effectiveness of national guidance in changing analgesic prescribing in primary care from 2002 to 2009: an observational database study. Eur J Pain 2013; 17(3):434–443.
39. Curtis HJ, Croker R, Walker AJ, et al. Opioid prescribing trends and geographical variation in England, 1998-2018: a retrospective database study. Lancet Psychiatry 2019;6(2):140–150.
40. Kurteva S, Abrahamowicz M, Gomes T, et al. Association of opioid consumption profiles after hospitalization with risk of adverse health care events. JAMA Netw Open 2021;4(5):e218782.
41. The Lancet Regional Health-Americas. Opioid crisis: addiction, overprescription, and insufficient primary prevention. Lancet Reg Health Am 2023;23:100557.
42. Macfarlane GJ, Barnish MS, Pathan E, et al. Co-occurrence and characteristics of patients with axial spondyloarthritis who meet criteria for fibromyalgia: results from a UK national register. Arthritis Rheumatol 2017;69(11):2144–2150.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 76–84
DOI 10.1002/acr.25659

Effect of Early Tumor Necrosis Factor Inhibitor Initiation on Chronic Opioid Use in Individuals With Axial Spondyloarthritis

Eva Petrow,[1] Jose A. Meade-Aguilar,[1,2] Christine Peloquin,[1] S. Reza Jafarzadeh,[1] Tuhina Neogi,[1] Maureen Dubreuil,[1,3] and Jean Liew[1]

Objective. Opioid use remains common despite effective therapies for axial spondyloarthritis (axSpA). We assessed whether early tumor necrosis factor inhibitor (TNFi) use is associated with reduced risk of chronic opioid use.

Methods. Using the Merative MarketScan Commercial Database containing health insurance billing claims data, we conducted a time-stratified, propensity score (PS)–matched cohort study of adults aged 18 to 65 years with axSpA. Early TNFi exposure was defined using pharmacy and medical claims as any incident TNFi use within 6 months of axSpA diagnosis. We used PSs to match early TNFi users 1:1 to those not starting a TNFi within 6 months in 1-year cohort accrual blocks. We compared the risk of chronic opioid use (≥90 days' prescription) among early TNFi users versus comparators using Cox proportional hazard models, overall and stratified by prior opioid use (within 12 months).

Results. We included 8,508 individuals with axSpA after PS matching (4,254 early TNFi initiators and 4,254 comparators) with a mean age of 42 years; 50% were female. Chronic opioid use occurred in 20.9% (22.3% early TNFi initiators and 19.6% comparators). Early TNFi initiators had 17% higher risk of chronic opioid use versus matched comparators (95% confidence interval [CI] 1.06–1.28), with higher risk of chronic opioid use for early TNFi initiators in the opioid-naive but not in the opioid-experienced stratum (hazard ratio [HR] 1.96; 95% CI 1.41–2.74 vs HR 1.06; 95% CI 0.96–1.17).

Conclusion. Early TNFi initiation was not associated with reduced risk of chronic opioid use versus later or no TNFi initiation. Confounding by indication in administrative claims data limit result interpretation.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a chronic inflammatory disease primarily affecting the axial skeleton, including the sacroiliac joints and spine, which causes significant pain and disability.[1] Although inflammation constitutes a major driver of pain in patients with axSpA, pain may be present in the absence of inflammation, is often multifactorial in origin, and may include bony damage or neuropathic pain.[2] First- and second-line treatments recommended by international guidelines for axSpA management focus on controlling inflammation, primarily through nonsteroidal anti-inflammatory drugs (NSAIDs) as initial pharmacotherapy, and tumor necrosis factor inhibitor (TNFi) or interleukin-17 inhibitor (IL-17i) therapy in patients who respond inadequately to NSAIDs.[3,4]

Not all patients with newly diagnosed axSpA achieve adequate symptom control with NSAIDs, and the presence of cardiac or kidney comorbidities limit their use.[5] A recent study showed that up to one-third of patients with axSpA continue to experience significant pain despite maximizing NSAID therapy.[6,7] Given this, it is not surprising that the chronic use of opioids among people with axSpA is common, with a prevalence of 23.5% when

Presented at the American College of Rheumatology Convergence 2024, Washington, DC, November 14–19, 2024.

Supported by the NIH (grant P30-AR-072571). Dr Neogi's work was supported by the NIH (grant K24-AR-070892). Dr Dubreuil's work was supported by the Arthritis Foundation, the Rheumatology Research Foundation, and the Boston University Clinical and Translational Science Institute. Dr Liew's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant K23-AR-082938); the Rheumatology Research Foundation Investigator Award; a Spondyloarthritis Research and Treatment Network pilot grant; and the Spondylitis Association of American Jane Bruckel Early Career Investigator Award.

[1]Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts; [2]Department of Medicine, Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts; [3]VA Boston Healthcare System, Boston, Massachusetts.

Drs Petrow and Meade-Aguilar are co-first authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25659).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25659.

Address correspondence via email to Jean Liew, MD, MS, at jwliew@bu.edu.

Submitted for publication April 29, 2025; accepted in revised form September 16, 2025.

SIGNIFICANCE & INNOVATIONS

- Individuals with axial spondyloarthritis (axSpA) may experience chronic pain despite guideline-recommended therapy with biologics or other targeted therapies, and chronic use of opioids among patients with axSpA is common. We hypothesized that early tumor necrosis factor inhibitor (TNFi) initiation would mitigate chronic opioid use.
- Using a US administrative claims database, we conducted a time-stratified, propensity score–matched cohort study of adults with axSpA to study the relationship of early TNFi initiation within 6 months of axSpA diagnosis (compared with later or no TNFi initiation) with the development of chronic opioid use. Contrary to our hypothesis, we did not find that early TNFi initiators had a reduced risk of chronic opioid use versus comparators.
- Confounding by indication limits interpretation of our results and may hinder other studies seeking to address the impact of biologic initiation on risk of chronic opioid use in this patient population. Additionally, more advanced study designs are needed to adequate account for dynamic treatment strategies.

identified using commercial claims data and a prevalence of 57.1% using Medicaid claims.[8] Relative to people with other rheumatic and musculoskeletal diseases, individuals with axSpA were nearly three times as likely to receive long-term opioid prescriptions.[9] Beyond the risk for developing opioid use disorder, chronic opioid use is also associated with increased multimorbidity,[10] health care utilization,[11] reduced quality of life, and death.[12]

Prior observational studies have identified risk factors associated with chronic opioid use among patients with axSpA, which include smoking, prior substance use disorder, a higher Charlson comorbidity index, depression, and anxiety.[10,13] Nonetheless, there is a paucity of recommendations directly addressing the use of opioid analgesics among individuals with axSpA. For example, opioid usage was not addressed by the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network treatment guidelines,[3] whereas the Assessment of SpondyloArthritis International Society/EULAR guidelines recommend cautious use of opioids and opioid-like drugs based on expert opinion.[4]

Therefore, a significant knowledge gap remains as to whether earlier initiation of current guideline-recommended therapies for axSpA, such as TNFis or other biologics, can mitigate chronic opioid use. It is plausible that earlier control of inflammation and disease activity may reduce the need for additional analgesic prescriptions. The aim of this study was to assess the potential protective role of early TNFi initiation among patients with axSpA for the development of chronic opioid use, using US administrative claims data.

PATIENTS AND METHODS

Data source and ethical approval. We used data from the Merative MarketScan Database, which consists of medical and drug claims data from employers and health plans.[14] It consists of data for >250 million individuals, including employees, their spouses, and dependents who are covered by employer-sponsored private health insurance in the United States. As this study involved the analysis of pre-existing, deidentified data, it was deemed exempt from requiring institutional review board (IRB) approval (Boston University Medical Campus IRB H-42058).

Study population. Claims were used to identify individuals with axSpA, ages 18 to 65 years old, from January 1, 2006, to December 31, 2021. AxSpA (including both radiographic and nonradiographic axSpA) was operationally defined as at least one inpatient claim with an axSpA *International Classification of Diseases* (ICD) *Ninth* or *Tenth Revision* diagnosis code or at least two outpatient claims ≥7 days apart with axSpA diagnosis codes. Such algorithms for identification of axSpA have previously been shown to have high positive predictive values (88%–100%), indicating their ability to correctly identify axSpA when the condition is present.[15]

Study design. We conducted a cohort study with the design shown in Figure 1. The date of study entry was the diagnosis date, defined as either the date of the first axSpA inpatient or outpatient claim among individuals meeting a definition of axSpA. Individuals with axSpA were observed until chronic opioid use (primary outcome), age 66 years, coverage end date, or the study end date (December 31, 2021). Follow-up ended after age 65 years as individuals become Medicare eligible at this age, and the dataset did not include services or prescriptions billed to Medicare.

Exposure of interest. The exposure of interest was any incident TNFi use within 6 months after the diagnosis date, reflecting early TNFi initiation, and was based on a combination of pharmacy claims and Current Procedural Terminology codes on medical claims for infusions (Supplementary Lists). As there is no consensus for early TNFi initiation in axSpA and early axSpA has been variably defined in the literature,[16] 6 months was chosen as adequate time to allow for TNFi initiation after axSpA diagnosis, but which would still be considered "early."

Outcome of interest. The primary outcome of interest was chronic opioid use, which was defined as any opioid use with duration of ≥90 days. A list of included opioid types and formulations is included in the Supplementary Lists. The index date for early TNFi initiators was the date of the first TNFi claim within 6 months of the diagnosis date (refer to Figure 1). For non-early TNFi initiators (comparators), the index date was a random date

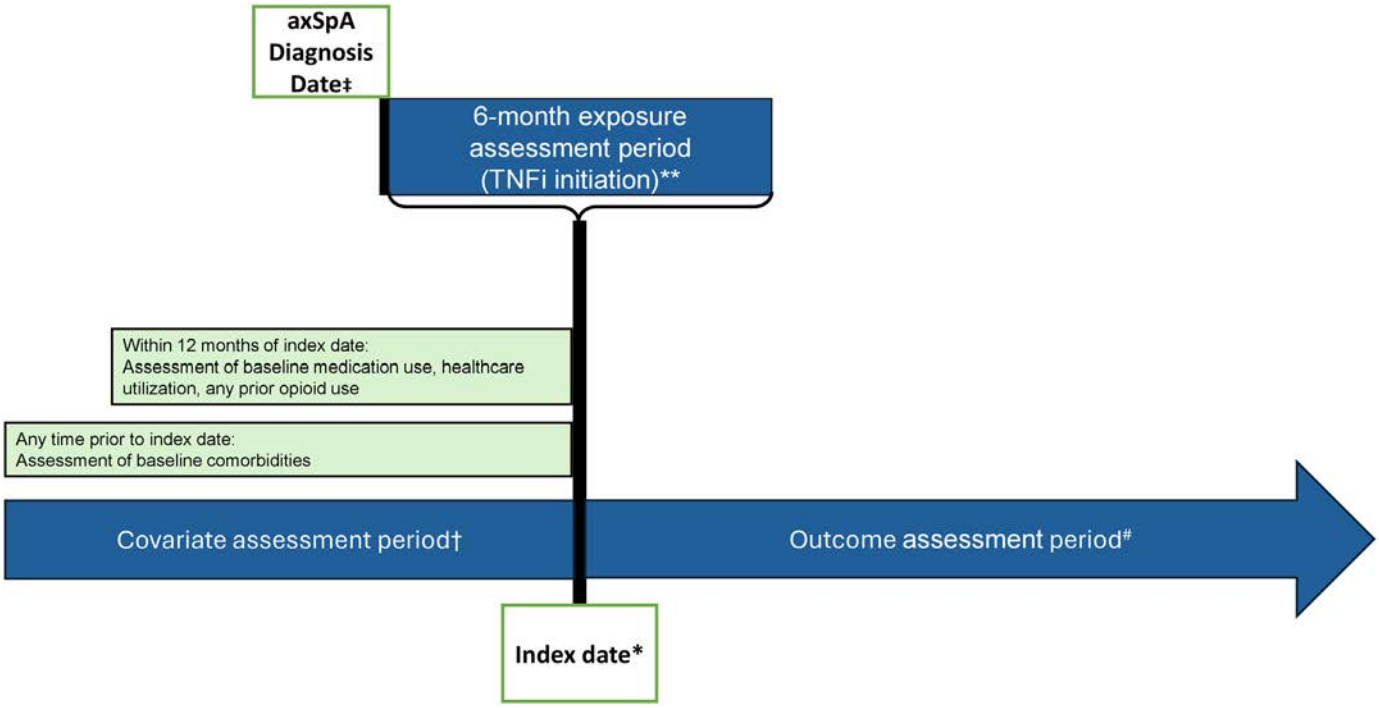


Figure 1. Cohort study design. ‡Assessment of eligibility criteria (2006–2021): age 18–65 years, at least 1 year of follow-up before the index date, axSpA diagnosis with at least one inpatient claim or at least two outpatient claims ≥7 days apart with *International Classification of Diseases Ninth* or *Tenth Revision* diagnosis codes. **TNFi initiation varies for each patient, occurring at any point during the 6-month exposure assessment period after the axSpA diagnosis date. †Covariates assessed include the following: age; sex; obesity; tobacco use; alcohol use disorder; cardiovascular disease; chronic obstructive pulmonary disease; chronic kidney disease; hypertension; diabetes (descriptive only); depression; use of conventional synthetic disease-modifying antirheumatic drugs; use of nonsteroidal anti-inflammatory drugs; use of glucocorticoids; use of interleukin-17 inhibitors, JAK inhibitors, and/or other non-TNFi biologics; any order for erythrocyte sedimentation rate and/or C-reactive protein; any primary care visit; any rheumatology outpatient visit; and any prior opioid use (for stratification). #Follow-up until outcome (chronic opioid use ≥90 days), coverage end date, 5 years of follow-up, or study end date (December 31, 2021). *Index date is the date of the first TNFi prescription for early TNFi initiators versus a random date within the 6-month exposure assessment period for non-TNFi initiators. axSpA, axial spondyloarthritis; TNFi, tumor necrosis factor inhibitor.

within 6 months of diagnosis date. To avoid overlap with the exposure assessment period and immortal time bias, the outcome assessment period started at the index date for all included patients (Figure 1).

Other covariates. Potential confounders for the relationship between early TNFi initiation and chronic opioid use were identified a priori. We constructed a directed acyclic graph to depict the presumed causal relationships and guide confounder adjustment (Supplementary Figure 1). To optimize the assessment of covariates, we included individuals needed to have 1 year of continuous enrollment before their diagnosis date. Covariates including comorbidities, alcohol use, obesity, and tobacco use were assessed from start of the continuous enrollment period, and covariates of medication use and health care utilization were assessed within 1 year before the index date (Figure 1), unless otherwise specified. Demographic confounders included age (assessed on the diagnosis date) and sex. Comorbidities associated with the exposure and/or the outcome included obesity, alcohol use disorder, tobacco use, cardiovascular disease, chronic kidney disease, chronic obstructive pulmonary disease, depression, and hypertension. A list of included covariate diagnosis and procedure codes is included in the Supplemental Lists. Medication use included any prescription of concomitant treatments for axSpA, including conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), NSAIDs, glucocorticoids, and IL-17is, JAK inhibitors, and/or other non-TNFi biologics. Surrogates of axSpA disease activity health care utilization included the presence of orders for erythrocyte sedimentation rate and/or C-reactive protein and history of any prior primary care and rheumatology outpatient visits.

Prior opioid use was also defined in the covariate assessment period. Opioid-experienced patients were defined by any opioid prescription within 1 year before the index date; otherwise, they were considered to be opioid-naive.

Statistical analysis. Variables were summarized using mean ± SD and proportions (%) as appropriate. We constructed a time-stratified, propensity score (PS)–matched cohort study to minimize confounding by indication and to account for secular

trends.[17,18] To calculate the PS, we used logistic regression with early TNFi use as the dependent variable, and we used the covariates described above, which reflect the indication for TNFi use and/or risk of chronic opioid use, as the independent variables. We matched early TNFi users to comparators 1:1 by the PS in 1-year cohort accrual blocks to account for secular trends using a greedy matching algorithm with nearest neighbor and within-caliper matching using a width of 0.1 SDs of the logit without replacement. A patient was included in a 1-year cohort accrual block if their index date was in the 1-year block. Covariate balance in the PS-matched dataset was assessed using the standardized mean difference (SMD) with an SMD <0.1 indicating a good covariate balance.

We compared the risk of chronic opioid use among PS-matched early TNFi initiators versus comparators using Cox proportional hazard models ("PS-matched model"). Additional Cox proportional hazard models were performed with further covariate adjustment: a "confounder-adjusted model" adjusted for confounders of age, sex, use of csDMARDs, use of glucocorticoids, use of IL-17is and/or JAK inhibitors and/or other biologics, use of NSAIDs, and the occurrence of any primary care and any rheumatology visits and a "PS-adjusted model" that included adjustment for all the PS model covariates in addition to confounders. The main model was the confounder-adjusted model. In addition to examining the cohort overall, we conducted a stratified analysis assessing opioid-experienced versus opioid-naive patients separately.

Sensitivity analyses. To assess the robustness of our results, we performed the following sensitivity analyses that included alternative axSpA definitions, different exposure assessment periods, and different definitions of chronic opioid use. These sensitivity analyses included the following: (1) a more lenient axSpA cohort definition with at least one inpatient or one outpatient ICD code; (2) a stricter axSpA definition that requires axSpA medication use either 6 months before or 6 months after the axSpA diagnosis code; (3) a longer exposure assessment period of 12 months; (4) a shorter exposure assessment period of 3 months; 5) a definition of chronic strong opioid use, which excludes low-potency opioids (tramadol, hydrocodone, and codeine); (6) a stricter definition of chronic opioid use defined as ≥120 days of opioid prescriptions; (7) a more lenient definition of chronic opioid use defined as ≥60 days of opioid prescriptions; (8) an outcome assessment start date with a lag of 3 months following the index date, as TNFi may take 3 months to take effect following initiation; and, lastly, (9) stratification by index date before or after January 1, 2014, to account for changes in opioid prescribing in the United States that occurred around the year 2014.

To assess potential bias from residual confounding, we estimated the effect size a potential unmeasured confounder would need to completely nullify the observed association between TNFi use and chronic opioid use in the main model through calculating an E-value.[19,20] A small E-value approaching 0 indicates that even a weak confounder may explain away the observed association, whereas a large E-value indicates that the observed effect is less likely because of confounding alone. To formally assess how the imperfect sensitivity and specificity of our algorithm to identify early TNFi initiation may have affected our effect measures, we performed a quantitative bias analysis for exposure misclassification. This estimated the hazard ratio (HR) for the primary outcome if there was misclassification in detecting early TNFi initiation by 10%, 20%, or 30%, corresponding to sensitivities of 90%, 80%, or 70%, respectively, while assuming 100% and 95% specificity (ie, that individuals without early TNFi initiation are not identified as being early initiators). These analyses were conducted using the R package "episensr."

RESULTS

In the main analysis, a total of 20,458 individuals were included (see Consolidated Standards of Reporting Trials flow diagram in Figure 2). Following PS matching, we included 8,508 individuals with axSpA, this included 4,254 early TNFi initiators and 4,254 non-early TNFi initiator comparators (Table 1). After PS matching, all SMDs were < 0.1 (Supplementary Table 1). Among the total propensity-matched cohort, the mean ± SD age was 42 ± 12 years, with 50% being female. Participants were well-matched in terms of frequencies of comorbidities, including 18% with obesity, 10% with tobacco use, 10% with cardiovascular disease, and 21% with depression. Other medication use in the overall, PS-matched cohort included 29% with csDMARD use and 66% with NSAID use. Half of the overall sample had any prior opioid use. The outcome of chronic opioid use was present in 20.9% of the overall study sample, and the median follow-up was 2 years. Overall, chronic opioid use was seen in 22.3% of early TNFi initiators and 19.6% of comparators.

In the main analysis of the PS-matched cohort, individuals starting TNFi early had a 17% higher risk of developing chronic opioid use compared with those who did not initiate TNFi early (HR 1.17; 95% confidence interval [CI] 1.06–1.28) (Table 2 and Figure 3) after additional confounder adjustment. The result of the PS-adjusted model was similar (HR 1.17; 95% CI 1.06–1.28) (Table 2).

Among the opioid-experienced, 38.2% of early TNFi initiators and 37.4% of comparators had chronic opioid use on follow-up (baseline demographics in Supplementary Tables 2A and B). In the opioid-naive group, chronic opioid use was observed in 5.0% of early TNFi initiators and 2.6% of comparators during follow-up. In the stratified analysis, after additional confounder adjustment we did not see an association between early TNFi initiation and chronic opioid use among opioid-experienced individuals (HR 1.06; 95% CI 0.96–1.17) (Table 2 and Figure 3). Among opioid-naive individuals, we observed a

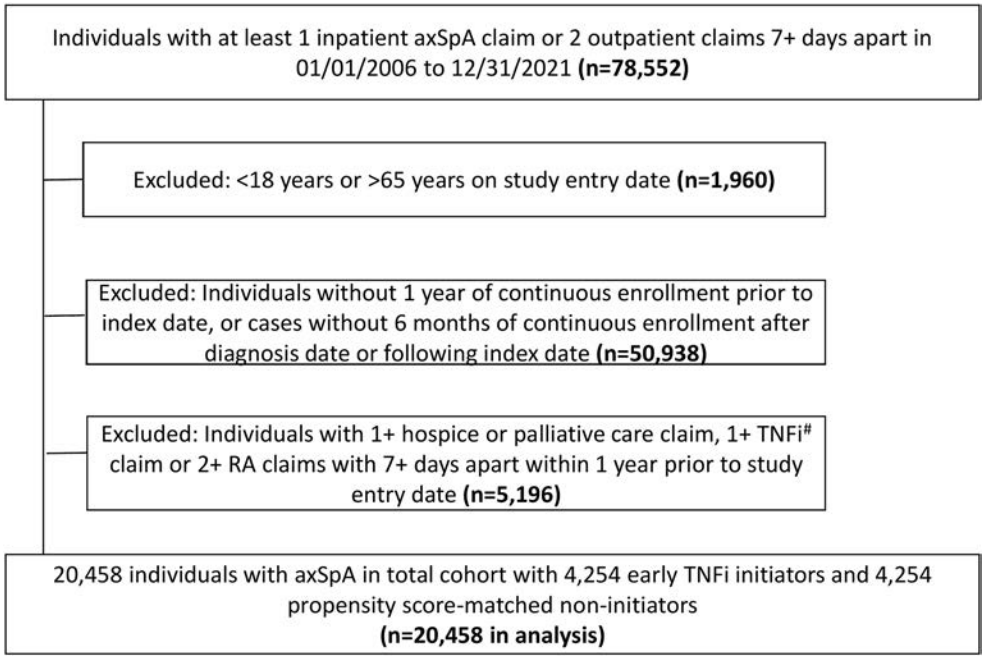


Figure 2. Consolidated Standards of Reporting Trials flow diagram for study inclusion and exclusion. #TNFis include etanercept, adalimumab, golimumab, certolizumab, and infliximab. axSpA, axial spondyloarthritis; RA, rheumatoid arthritis; TNFi, tumor necrosis factor inhibitor.

higher risk of chronic opioid use for those initiating TNFi early compared with those who did not (HR 1.96; 95% CI 1.41–2.74) (Table 2 and Figure 3).

Within the sensitivity analyses (Table 2), we observed similar results to the main analysis, regardless of the axSpA cohort definition, exposure assessment period, or outcome definition. Among

Table 1. Demographics and baseline clinical characteristics in the propensity score–matched cohort, overall and for exposed and unexposed*

Covariate	Overall (N = 8,508)	Early TNFi initiators (n = 4,254)	Nonearly TNFi initiators (n = 4,254)
Age, mean ± SD, y	42.3 ± 11.8	42.4 ± 11.6	42.3 ± 11.9
Female sex, %	50.5	50.1	51.0
Obesity, %	17.8	18.1	17.5
Alcohol use disorder, %	2.6	2.9	2.4
Tobacco use, %	9.9	10.3	9.4
CVD, %	10.0	10.2	9.9
CKD, %	7.0	6.9	7.0
COPD, %	9.2	9.3	9.0
Diabetes, %	12.3	11.7	12.9
Depression, %	21.1	21.0	21.1
Hypertension, %	33.1	33.9	32.3
Prior opioid use (any),[a] %	50.1	50.9	49.3
csDMARD use, %	28.7	28.7	28.8
NSAID use, %	65.8	65.7	65.9
Glucocorticoid use, %	48.8	49.6	48.1
IL-17i/JAKi/other biologic use,[b] %	0.40	0.40	0.40
ESR/CRP testing orders (any), %	89.3	88.6	89.9
Primary care visit (any),[c] %	86.0	86.0	86.1
Rheumatology visit (any),[c] %	72.4	72.0	72.8

* CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; csDMARD, conventional synthetic disease-modifying antirheumatic drug; CVD, cardiovascular disease, ESR, erythrocyte sedimentation rate; IL-17i, interleukin-17 inhibitor; JAKi, JAK inhibitor; NSAID, nonsteroidal anti-inflammatory drug; TNFi, tumor necrosis factor inhibitor.
[a] Defined as any opioid prescription claim within covariate assessment period 12 months before the index date.
[b] csDMARDs, NSAIDs, glucocorticoids, IL-17i JAKis, and other biologics lists are at the end of the Supplementary Lists.
[c] Defined as any primary care visit or any rheumatology visit within the covariate assessment period 12 months before the index date.

Table 2. Results from multivariable analyses assessing the relation of early tumor necrosis factor inhibitor initiation with the risk of chronic opioid use among individuals with axSpA: main, stratified, and sensitivity analyses*

	PS-matched HR (95% CI)	Confounder-adjusted HR (95% CI)[a]	PS-adjusted HR (95% CI)[b]
Main analysis			
≥ 90 days opioid use	1.16 (1.06–1.27)	1.17 (1.06–1.28)	1.17 (1.06–1.28)
Stratified analyses			
Opioid-experienced	1.05 (0.95–1.16)	1.06 (0.96–1.17)	1.07 (0.97–1.18)
Opioid-naive	1.96 (1.41–2.73)	1.96 (1.41–2.74)	1.99 (1.43–2.78)
Sensitivity analyses			
Lenient axSpA definition[c]	1.21 (1.11–1.33)	1.23 (1.13–1.35)	1.21 (1.11–1.33)
Strict axSpA definition[d]	1.11 (0.96–1.28)	1.11 (0.96–1.28)	1.16 (1.01–1.35)
12-month exposure assessment period	1.31 (1.19–1.45)	1.31 (1.19–1.45)	1.34 (1.21–1.47)
3-month exposure assessment period	1.12 (1.01–1.24)	1.13 (1.02–1.25)	1.12 (1.01–1.24)
Chronic strong opioid use only[e]	1.21 (1.07–1.35)	1.21 (1.08–1.36)	1.20 (1.07–1.35)
≥ 120 days opioid use	1.15 (1.04–1.27)	1.16 (1.05–1.27)	1.15 (1.05–1.27)
≥ 60 days opioid use	1.18 (1.08–1.29)	1.19 (1.09–1.29)	1.19 (1.09–1.30)
Outcome assessment with 3-month lag	1.17 (1.06–1.29)	1.17 (1.06–1.29)	1.16 (1.05–1.28)
Index date 2007–2013	1.17 (1.04–1.32)	1.18 (1.04–1.33)	1.19 (1.05–1.34)
Index date 2014–2021[f]	1.14 (0.99–1.32)	1.15 (0.99–1.33)	1.13 (0.98–1.31)

* axSpA, axial spondyloarthritis; CI, confidence interval; HR, hazard ratio; PS, propensity score.
[a] Adjusted for all confounders: age; sex; use of conventional synthetic disease-modifying antirheumatic drugs; use of glucocorticoids; use of interleukin-17 inhibitors, JAK inhibitors, and/or other biologics; use of nonsteroidal anti-inflammatory drugs; and any primary care and rheumatology visits.
[b] Adjusted for all PS covariates, including those that were not true confounders: age; sex; obesity; tobacco use; alcohol use disorder; cardiovascular disease; chronic obstructive pulmonary disease; chronic kidney disease; hypertension; depression; use of conventional synthetic disease-modifying antirheumatic drugs; use of glucocorticoids; use of interleukin-17 inhibitors, JAK inhibitors, and/or other biologics; use of nonsteroidal anti-inflammatory drugs; any erythrocyte sedimentation rate/C-reactive protein laboratory orders; and any primary care and rheumatology visits.
[c] Defined as at least one inpatient or at least one outpatient axSpA diagnostic code.
[d] Defined as at least one inpatient or at least two outpatient axSpA diagnostic codes ≥7 days apart plus concomitant disease-modifying antirheumatic drug use with any prescription within 6 months before or after the first diagnosis code of any of the following: tumor necrosis factor inhibitors, interleukin-17 inhibitors, JAK inhibitors, conventional synthetic disease-modifying antirheumatic drugs, or other biologics.
[e] Chronic opioid use was defined as >90 days, which excludes low-potency opioid types including codeine, tramadol, and hydrocodone.
[f] Stratified by the index date before or after January 1, 2014.

these, the strongest association was observed with the use of a 12-month exposure assessment period, and individuals starting TNFi early had a 31% higher risk of developing chronic opioid use compared with those who did not initiate TNFi early (HR 1.31; 95% CI 1.19–1.45). This association was weaker with the 3-month exposure assessment period analysis (HR 1.13; 95% CI 1.02–1.25). Otherwise, the risk of chronic opioid use among early TNFi initiators compared with noninitiators did not differ from the main analysis when using a cohort definition based on a lenient definition of axSpA with at least one inpatient or one outpatient ICD code (HR 1.23; 95% CI 1.13–1.35 with confounder adjustment). No association was seen between early TNFi initiation and chronic opioid use among opioid-experienced individuals when using a stricter axSpA definition that included axSpA medication use (HR 1.11; 95% CI 0.96–1.28). Further, similar results were seen between each of the chronic opioid use definitions, including a strong chronic opioid use definition excluding low-potency opioids (HR 1.21; 95% CI 1.08–1.36), a stricter definition (≥120 days of prescriptions) of chronic opioid use (HR 1.16; 95% CI 1.05–1.27), and a more lenient definition (≥60 days of prescriptions) (HR 1.19; 95% CI 1.09–1.29). Lastly, the results did not differ when using an outcome assessment start date with a lag of 3 months following the index date (HR 1.17; 95% CI 1.06–1.29) nor with stratification by index date before or after 2014 (HR 1.18; 95% CI 1.04–1.33 vs HR 1.15; 95% CI 0.99–1.33, respectively).

In the E-value analysis, our observed HR of 1.17 for the main results could be fully explained by an unmeasured confounder associated with both the exposure and the outcome by a HR of 1.47 above and beyond the measured confounders. The results of the quantitative bias analysis, assuming a sizable 30% exposure misclassification (corresponding to 70% sensitivity of our algorithms for early TNFi initiation), showed that the HR for the primary outcome would be 1.20 (95% CI 1.07–1.35), similar to our observed PS-matched HR of 1.17 for early TNFi initiators versus comparators. The results were similar when assuming 20% and 10% exposure misclassification and when specificity was set at 95% (Supplementary Table 3).

DISCUSSION

In this large cohort study of individuals with axSpA using US administrative claims data, we investigated the association

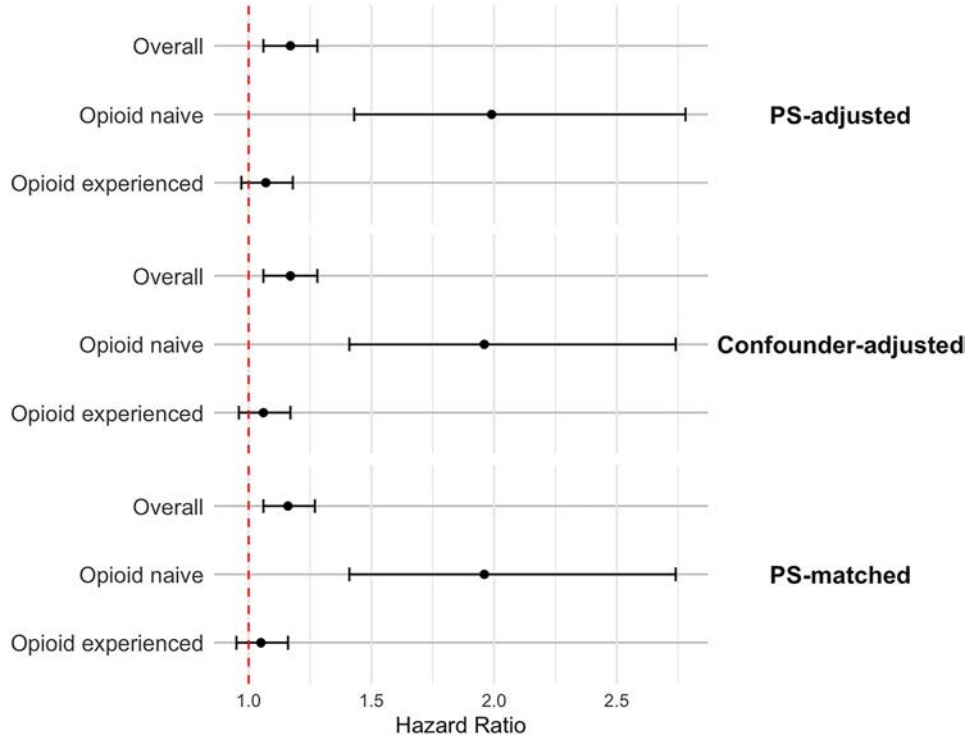


Figure 3. Association of early tumor necrosis factor inhibitor use compared with nonearly tumor necrosis factor inhibitor use for the outcome of chronic opioid use, in the overall PS-matched study cohort and stratified by prior opioid use. Results from PS-matched, confounder-adjusted, and PS-adjusted Cox models. PS, propensity score. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25659/abstract.

between early initiation of TNFi therapy and subsequent chronic opioid use. Contrary to our initial hypothesis, we did not find evidence that early TNFi initiation was associated with a reduced risk of chronic opioid use. Among opioid-naive individuals, early TNFi use was surprisingly associated with a higher likelihood of chronic opioid use, whereas no significant association was observed among opioid-experienced individuals. However, our results should be interpreted with caution, given the limitations inherent to claims data related to confounding by indication whereby unmeasured factors are likely to impact both early TNFi initiation and chronic opioid prescription patterns in this population.

Chronic opioid use, known to be associated with increased risk of fractures, higher health care utilization, and more frequent hospitalizations,[13] represents a significant health problem in patients with axSpA. Our results align with previous studies regarding the burden of chronic opioid use among individuals with axSpA, highlighting the high prevalence of chronic opioid use in this population, despite the rising availability of effective biologic therapies including TNFi in the past two decades. We found that chronic opioid use was present in 20.9% of the overall study sample. A Finnish cohort study showed that, after 1 year of diagnosis, patients with axSpA had a cumulative incidence of chronic opioid use of 21.7%, whereas the incidence in matched comparators without axSpA was 7.8%.[21] However, our prevalence of chronic opioid use was lower than that previously estimated among Medicaid covered individuals (57.1% in the study by Sloan et al[8]), suggesting that public insurance may pose additional risks for chronic opioid use that warrant further investigation.

Prior research efforts have focused on elucidating contributing and protective factors for chronic opioid use in axSpA. A UK cohort study identified several key social and health factors—such as active smoking, concurrent substance use disorder, presence of social deprivation factors, suicidality, fibromyalgia, and higher multimorbidity—as significant predictors of chronic opioid use.[13] Similarly, Dau et al identified factors such as longer disease duration, smoking, and higher disease activity as being associated with opioid use in a prospective cohort of patients with radiographic axSpA (ankylosing spondylitis), but there was a null association between TNFi use and opioid use.[10] It remains unclear whether treatment strategies such as early initiation of biologic or second-line therapies can prevent opioid dependence in this population.

This study is the first to examine the impact of early TNFi therapy, initiated shortly after axSpA diagnosis, on the risk of developing chronic opioid use. Our study adds to the literature suggesting that early TNFi initiation alone does not appear to mitigate opioid use. This aligns with the population-based study from the United Kingdom assessing the development of incident chronic opioid use, which included patients with both axSpA and psoriatic arthritis.[13] Notably, this study found that none of the guideline-recommended therapies for axSpA or psoriatic arthritis were linked to a lower likelihood of chronic opioid use.

The unexpected finding that early TNFi initiation was linked with a higher risk of chronic opioid use, especially in opioid-naive individuals, warrants further consideration of potential mechanisms driving this association, including methodologically through confounding by indication. This type of unmeasured or residual confounding occurs when patients who begin TNFi therapy earlier in their disease course may have more active, severe, or refractory disease, resulting in persistent inflammation and pain,[22] leading to long-term opioid use despite appropriate biologic intervention.

The presence of residual confounding is supported by a low calculated E-value of 1.47, indicating that an unmeasured confounder would need to be associated with both the treatment and the outcome with an HR of ≥1.47 to fully explain away the association. It is possible that a combination of unmeasured or residual confounding may have produced this E-value. This hypothesis is further supported by previous studies that have identified a subset of patients with axSpA who continue to experience significant pain despite receiving optimal treatment.[6] Furthermore, individuals requiring chronic opioid use therapy may also exhibit higher health care utilization, which could influence early TNFi initiation.[11] To address this potential confounding, we included available health care utilization variables from our claims database into PSs and further adjusted by these variables after PS matching to strengthen the validity of our findings. However, we were unable to account for other important drivers of chronic opioid use, such as disease activity and specific patient characteristics including social determinants of health, which could not be measured in our study. The available confounders extractable from administrative claims data like MarketScan may not be sufficient to fully address confounding by indication for our primary research question of whether early TNFi initiation reduces the risk of chronic opioid use in individuals with axSpA.

Beyond confounding by indication, misclassification for the exposure of early TNFi initiation is another potential threat to the validity of our results. This misclassification may arise because some individuals with axSpA may have previously been exposed to TNFi, but this information is not captured in claims data. To address this possible exposure misclassification, we performed quantitative bias analysis, which demonstrated insubstantial changes to our observed results, suggesting against exposure misclassification as a major source of bias in our study.

Although PS methods are valuable tools for balancing measured confounders in observational studies, they may not fully capture the complexity of clinical decision-making and changing treatment patterns over time. In our setting, both TNFi initiation and opioid therapy duration are time-dependent, yet traditional PS approaches are unable to fully account for treatment changes over follow-up. As such, our estimates may be biased by these factors. More advanced causal inference methods would be better suited to address these complexities and more accurately estimate the causal effects in the presence of dynamic treatment strategies. However, such methods would still be limited by the presence of unmeasured confounders.

Further limitations of this study should be acknowledged. First, as with all administrative claims–based studies, there may be misclassification in defining the axSpA population, including the possibility of missed axSpA diagnoses. Although axSpA diagnostic claims marked the start of the 6-month exposure assessment period, it is possible that patients may have had longstanding prevalent disease, creating a heterogenous study population with varying disease duration. Second, as described above, residual confounding may persist despite our rigorous study design using PS methods to account for confounding, as patient-reported pain levels, disease severity or disease activity, and functional impairment were not directly measurable using claims data. Race and ethnicity were not available in our dataset and represent further potential unmeasured confounding. Third, indications for chronic opioid use were not available, and opioids may have been prescribed for reasons related to persistent axSpA symptoms or for other reasons. Lastly, our exposure was assessed at the beginning of the study, before follow-up, and we did not allow for dynamic assessment of the exposure, meaning that we could not account for individuals who started TNFi later during the study period. This could introduce bias to our results, as patients who never received TNFi may significantly differ from patients who received TNFi treatment later in their disease course, yet both types of individuals were considered together in the comparator group. Our study has several strengths, including the use of a large, national administrative claims database, a robust PS-matching approach to minimize confounding, and multiple sensitivity analyses to test the robustness of our findings.

In conclusion, our study provides additional evidence that early TNFi initiation appears not to be associated with a lower risk of chronic opioid use in individuals with axSpA. Although our results suggest a connection between early TNFi initiation and higher risk of chronic opioid use in patients without prior exposure to opioids, confounding by indication limits this interpretation. For future studies to better clarify the role of early TNFi therapy in terms of opioid-sparing effects among individuals with axSpA, the ideal data source should include important confounders such as disease severity, disease activity, and measures of pain. Advanced methods that can better account for dynamic treatment changes may also mitigate threats to validity. In the meantime, clinicians should continue to implement multimodal treatment strategies to reduce opioid reliance among individuals with axSpA beyond the initiation of biologic therapy.

ACKNOWLEDGMENT

We thank Maggie Westerland for performing the quantitative bias analysis for exposure misclassification.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Taurog JD, Chhabra A, Colbert RA. Ankylosing spondylitis and axial spondyloarthritis. N Engl J Med 2016;375(13):1303.
2. Bittar M, Deodhar A. Axial spondyloarthritis: a review. JAMA 2025; 333(5):408–420.
3. Ward MM, Deodhar A, Gensler LS, et al. 2019 update of the American College of Rheumatology/Spondylitis Association of America/ Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. Arthritis Care Res (Hoboken) 2019;71(10): 1285–1299.
4. Ramiro S, Nikiphorou E, Sepriano A, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. Ann Rheum Dis 2023;82(1):19–34.
5. Webers C, Boonen A, Vonkeman HE, et al. Factors associated with residual disease in axial spondyloarthritis: results from a clinical practice registry. J Rheumatol 2023;50(11):1430–1438.
6. Philippoteaux C, Delepine T, Cailliau E, et al. Characteristics of difficult-to-treat axial spondyloarthritis: results of a real-world multicentric study. Joint Bone Spine 2024;91(2):105670.
7. Sieper J, Lenaerts J, Wollenhaupt J, et al; All INFAST Investigators. Efficacy and safety of infliximab plus naproxen versus naproxen alone in patients with early, active axial spondyloarthritis: results from the double-blind, placebo-controlled INFAST study, part 1. Ann Rheum Dis 2014;73(1):101–107.
8. Sloan VS, Sheahan A, Stark JL, et al. Opioid use in patients with ankylosing spondylitis is common in the United States: outcomes of a retrospective cohort study. J Rheumatol 2019;46(11):1450–1457.
9. Chen SK, Feldman CH, Brill G, et al. Use of prescription opioids among patients with rheumatic diseases compared to patients with hypertension in the USA: a retrospective cohort study. BMJ Open 2019;9(6):e027495.
10. Dau JD, Lee M, Ward MM, et al. Opioid analgesic use in patients with ankylosing spondylitis: an analysis of the prospective study of outcomes in an ankylosing spondylitis cohort. J Rheumatol 2018;45(2): 188–194.
11. Liew JW, Pedro S, Michaud K, et al. Opioid use and healthcare utilization in adults with psoriatic arthritis and ankylosing spondylitis. J Rheumatol 2024;51(10):1009–1014.
12. Hayes CJ, Payakachat N, Li C. Evaluation of opioid use among patients with back disorders and arthritis. Qual Life Res 2018;27(11): 3021–3035.
13. Huang YT, Jenkins DA, Yimer BB, et al. Factors associated with long-term opioid use among patients with axial spondyloarthritis or psoriatic arthritis who initiated opioids. Rheumatology (Oxford) 2025; 64(4):1844–1852.
14. IBM MarketScan Research Databases for life sciences researchers. Accessed August 1, 2024. https://www.merative.com/documents/brief/marketscan-explainer-general
15. Dubreuil M, Peloquin C, Zhang Y, et al. Validity of ankylosing spondylitis diagnoses in The Health Improvement Network. Pharmacoepidemiol Drug Saf 2016;25(4):399–404.
16. Benavent D, Capelusnik D, van der Heijde D, et al. How is early spondyloarthritis defined in the literature? Results from a systematic review. Semin Arthritis Rheum 2022;55:152032.
17. Dubreuil M, Zhu Y, Zhang Y, et al. Allopurinol initiation and all-cause mortality in the general population. Ann Rheum Dis 2015;74(7): 1368–1372.
18. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. Pharmacoepidemiol Drug Saf 2005;14(7):465–476.
19. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. JAMA 2019;321(6):602–603.
20. Mathur MB, Ding P, Riddell CA, et al. Web site and R package for computing E-values. Epidemiology 2018;29(5):e45–e47.
21. Muilu P, Rantalaiho V, Kautiainen H, et al. Opioid use frequency in early axial spondyloarthritis in Finland - a pharmacoepidemic register study. Joint Bone Spine 2022;89(3):105302.
22. Glintborg B, Sørensen IJ, Østergaard M, 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.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 85–95
DOI 10.1002/acr.25679

American College of Rheumatology
Empowering Rheumatology Professionals

Digital Behavioral Therapy Improves Outcome in Patients With Axial Spondyloarthritis and Persistent Pain: A Randomized Controlled Trial

David Kiefer,[1,2,5] Yade Sonkaya,[1,2] Dietmar Krause,[1,3] Markus Voglau,[4] Bernhard Mintrop,[1,5] Imke Redeker,[1,2] Xenofon Baraliakos,[1,2] and Uta Kiltz[1,2]

Objective. Axial spondyloarthritis (axSpA) is often associated with persistent pain despite effective anti-inflammatory treatment. Digital health applications (DHAs) provide innovative approaches to address multidimensional aspects of persistent pain through psychological and behavioral strategies. The aim of this study was to assess the impact of a DHA using acceptance and commitment therapy (ACT) on disease outcomes, including the West Haven-Yale Multidimensional Pain Inventory (MPI), in patients with axSpA experiencing persistent pain despite stable pharmacological therapy.

Methods. This unblinded, multicentric, randomized controlled trial compared an intervention group (IG) receiving the ACT app with a standard of care (SOC) group. The ACT app provided behavioral therapy. The primary outcome was MPI pain-related life interference; secondary outcomes included pain severity, affective distress, and other patient-reported outcomes after 12 weeks. Linear models estimating the effect of the ACT app on the change of MPI pain-related life interference and affective distress were calculated.

Results. A total of 136 patients were randomized to IG (n = 73) with the ACT app and SOC (n = 63) without the ACT app. In the IG, 44 actively used the ACT app. All lessons in the ACT app were completed by 19 IG patients (43%). Baseline characteristics, including MPI scores, were comparable between groups. IG showed a reduction in pain-related life interference as well as in other outcomes. The improvements in pain-related life interference (β with −0.36, 95% confidence interval [CI]: −0.73 to 0.01) and affective distress related to the disease (−0.4; 95% CI −0.84 to 0.03) were greater compared with SOC.

Conclusion. The ACT app demonstrated a meaningful reduction in pain-related life interference, supporting that DHAs might become a complementary tool in managing pain for patients with axSpA. Studies about improving adherence to DHAs are warranted.

INTRODUCTION

Axial spondyloarthritis (axSpA) is a chronic inflammatory disease primarily affecting the axial skeleton, often leading to pain, stiffness, and functional limitations.[1] Over the past decades, advancements in pharmacologic therapies, including biologic (b) and targeted synthetic (ts) disease-modifying antirheumatic drugs (DMARDs) targeting inflammatory pathways, have significantly improved the outcome of patients.[2] However, a notable subset of patients with axSpA continues to report persistent pain despite effective anti-inflammatory treatment.[3] As a frequently occurring symptom reported in patients with axSpA, persistent pain is associated with lower health-related quality of life (HR-QoL), fatigue, and functional and work productivity impairments.[4,5] Therefore, this persistent pain presents a considerable challenge in clinical practice. The typical inflammatory back pain is a key symptom of patients with axSpA.[6] However, the presence of back pain may reflect disease activity but can also potentially be caused by secondary components, such as coping mechanisms, as well as physical and psychological comorbidities.[7]

Pain management in patients with axSpA primarily involves pharmacotherapy combined with nonpharmacologic treatments,

Supported by Ruhr-Universität Bochum, Germany (Forum grant F1085-2023).

[1]Ruhr-Universität Bochum, Bochum, Germany; [2]Rheumazentrum Ruhrgebiet, Herne, Germany; [3]Private rheumatology practice, Gladbeck, Germany; [4]Private rheumatology practice, Oldenburg, Germany; [5]Private rheumatology practice, Hattingen, Germany.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25679.

Address correspondence via email to Uta Kiltz, MD, at uta.kiltz@elisabethgruppe.de

Submitted for publication April 30, 2025; accepted in revised form September 22, 2025.

SIGNIFICANCE & INNOVATIONS

- Digital health applications may have the potential to decrease the impact of pain on daily life for patients with axial spondyloarthritis.
- Digital interventions based on acceptance and commitment therapy may improve health-related quality of life for those with axial spondyloarthritis.
- Digital health applications might serve as a useful addition to standard care in pain management.
- Technical feasibility and usability ratings for the app were generally positive, but long-term adherence remains a challenge.

such as exercise, psychological interventions, and self-management support.[2] Particularly in patients with predominantly nociplastic and somatoform pain components (eg, fibromyalgia), behavioral therapy approaches play a central role. Various clinical studies have demonstrated that within the framework of multimodal treatment concepts, behavioral therapy measures constitute a key pillar of this approach.[8] This is also reflected in the current recommendations by EULAR for fibromyalgia, which endorse cognitive behavioral therapy (CBT) as an effective treatment.[9] Initial approaches to digital solutions have already been investigated and published in clinical studies.[10] Furthermore, the use of electronic health apps has been shown to effectively reduce pain intensity and its associated interference for patients with chronic back pain.[11] These applications often integrate evidence-based psychological therapy and CBT, such as acceptance and commitment therapy (ACT), to address the complex interplay of physical and psychological factors contributing to pain.[12] By fostering the acceptance of pain, enhancing psychological flexibility, and promoting engagement in meaningful activities, digital health applications (DHAs) hold the potential to improve multidimensional pain outcomes and HR-QoL.[13,14]

In Germany, DHAs are being reimbursed by statutory health insurance if they have been approved by the German Federal Institute for Drugs and Medical Devices (BfArM).[15] In March 2025, 68 DHAs in total have been evaluated in terms of efficacy and safety by the regulatory authority. The prescription of the DHA follows a standardized way, which includes an approval step of the health insurance provider, followed by the provision of an activation code to access the DHA.[15] The costs are fully covered by the patient's health insurance. This system represents the first worldwide opportunity to prescribe DHAs in health care in a manner akin to medical devices or therapeutic aids.

So far, within the field of musculoskeletal disorders, DHAs have only been approved for mental health conditions or orthopedic disorders, reflecting the primary focus of digital therapeutics to date. However, no DHA has been specifically approved for inflammatory and rheumatic and musculoskeletal diseases (RMDs), including axSpA.[15] Given that pain management and ACT play an important role in the management of axSpA, DHAs have a significant potential for their application in this field. The aim of this study was to assess the impact of a DHA using the ACT on disease outcomes in patients with axSpA experiencing persistent pain despite stable pharmacologic therapy.

PATIENTS AND METHODS

Study design. In this multicentric, unblinded, randomized controlled trial, adult patients with a clinical diagnosis of axSpA were eligible for inclusion. This trial was designed with pragmatic elements, aiming to reflect routine clinical practice conditions. Inclusion criteria required a symptom duration of at least 6 months, stable disease control under pharmacologic therapy, and a pain intensity ≥4 on a numerical rating scale (NRS; 0–10, 10 = severe pain) at randomization. Stable disease control was defined as the treating rheumatologist's judgment that no change in immunomodulatory or analgesic therapy was indicated, even if disease activity scores were elevated because of extreme high patient-reported outcome (PRO) results. This pragmatic approach was chosen to reflect real-world clinical practice and to target the patient population most likely to benefit from a DHA. Exclusion criteria included language barriers, inability to provide informed consent, planned changes to or conventional or bDMARD therapy, or an inability or unwillingness to use DHAs.

Patients were randomized 1:1 to either an intervention group (IG) or a standard of care (SOC) group and were observed over 12 weeks. Randomization was conducted using a table with random numbers. Only personnel who assigned participants to the intervention had access to the random allocation sequence, which was not the case for personnel who enrolled participants. SOC patients continued to receive treatment as determined by their treating rheumatologist, without the use of the ACT app. IG patients who did not activate the ACT app within 4 weeks received a reminder phone call from the study team to help with obtaining and using the activation code for the DHA. Baseline assessments were repeated if patients activated the DHA 4 weeks after the initial baseline assessment. We defined active app users as those who engaged with the app at least once, and inactive app users as those who either did not use the app at all or were unable to access it because of insurance-related issues.

DHA. We used the HelloBetter Ratiopharm Chronischer Schmerz (ACT app), which is an interactive psychological online program designed to sustainably reduce the impact of persistent pain and is approved as a DHA for persistent somatoform pain disorder, persistent pain disorder with somatic and psychological factors, and fibromyalgia.[15,16] Through evidence-based psychoeducation provided via texts, videos, and audio content, participants learn effective strategies from ACT.[15,17] These strategies are reinforced through practical exercises that can be integrated into daily life, helping individuals improve their management of persistent pain and reduce its impact on their QoL.

The HelloBetter app online program is aimed at the long-term reduction of pain-related life interference and is available for PCs, tablets, or smartphones. The program consists of seven units, each requiring 45 to 60 minutes to complete. The content was developed by psychologists and psychotherapists with expertise in persistent pain management and digital health, in accordance with established ACT protocols. Its efficacy and safety were evaluated in clinical trials, and it is listed as an approved DHA in the official directory of the BfArM.[15,16,18,19]

Assessment of axSpA. Data on demographic and disease status as well as information on laboratory results (eg, C-reactive protein [CRP] and HLA-B27) and data on current treatment including the nonsteroidal anti-inflammatory drug (NSAID) usage indices were taken from medical records. Patients responded to a standardized tool of PROs at baseline and after 3 months. Pain assessment included a global pain question (NRS 0–10) and the PainDETECT questionnaire, which ranged from 1 to 24, with a threshold of >18 indicating probable neuropathic pain.[20] Furthermore, the Multidimensional Pain Inventory (MPI) was assessed.[21] MPI Part 1, "Pain Experience and Psychosocial Impact," evaluates the impact of pain across five dimensions using 22 questions:

1. Pain severity: the intensity of the pain experienced.
2. Interference: the extent to which pain disrupts daily life, activities, and overall functioning.
3. Perceived support: the level of emotional and practical support from others.
4. Life control: the degree to which patients feel in control of their pain and life circumstances.
5. Affective distress: the emotional burden associated with living with chronic pain.

Each MPI question is scored on a 7-point Likert scale (0–6), and subscores are calculated by averaging the responses within each dimension.[21]

Disease activity was measured with the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Axial Spondyloarthritis Disease Activity Score (ASDAS), which both primarily target nociceptive pain associated with inflammatory activity. Furthermore, disease-specific assessments included the Bath Ankylosing Spondylitis Functional Index (BASFI), the Physician Global Assessment (PhGA), and the Assessment of Spondyloarthritis International Society Health Index (ASAS-HI). The Hospital Anxiety and Depression Scale (HADS) was used to evaluate psychological comorbidities. The Fibromyalgia Impact Questionnaire (FIQ) was assessed for pain related to fibromyalgia. The mental component score (MCS) and physical component score (PCS) of the 36-item Short Form were used to assess physical and mental HR-QoL. Physical activity levels were quantified using the modified Short Questionnaire to Assess Health-enhancing physical activity (mSQUASH).

Assessment of usage of the DHA. Usage of the DHA was provided by self-report and electronic tools. IG patients provided self-reports of app usage, detailing the number of completed modules. This was complemented by electronically recorded data to objectively assess engagement, including the percentage of the program completed and the time spent using the app. The mobile app rating scale (MARS) was used to assess app engagement, functionality, aesthetics, and subjective quality, along with a system usability score to quantify the overall user experience. The evaluation resulted in an average value for each section and an average value for the overall app quality between 1 and 5. A value of ≥3 was considered moderate and a value of ≥4 was considered good.[22] Acceptance of the DHA was assessed using the Net Promoter Score (NRS 0–10): patients with a score from 0 to 6 are considered to be detractors, patients with a score from 7 to 8 are considered to be passives, and patients with a score ≥9 are considered to be promoters.[23] The Patient Global Impression of Change (PGIC) was used to evaluate patients' subjective perception of overall change during the study. Patients rate a single-item questionnaire their perception of change on a 7-point Likert scale, ranging from "very much improved" (1) to "very much worse" (7).[24] To identify potential barriers to using the app, we developed a 10-item study-specific questionnaire addressing aspects such as usability, technical problems, user adherence, individual needs, data protection, doctor-patient relationship, digital expertise, implementation into therapy, and costs. Patients rated each item on a scale from 0 (no barrier) to 10 (significant barrier). The System Usability Score is a validated 10-item questionnaire that evaluates the usability of a system, providing a single score ranging from 0 to 100, with higher scores indicating better usability.[25]

The primary endpoint of the study was change in the West Haven-Yale MPI dimension of pain-related life interference after 12 weeks. Secondary endpoints included axSpA PRO measurements and the other dimensions assessed through the MPI.[21]

Statistical analyses. Differences between baseline and follow-up were examined using paired *t*-tests for continuous variables and the McNemar chi-square test for categorical ones. Baseline adjusted differences in pain-related interference and emotional distress between active and inactive app users were analyzed using analysis of covariance. A P value of <0.05 was considered to be significant.

We assumed for the sample size calculation a sample variance of 1.162 for the MPI after 3 months, a correlation coefficient of 0.5 between pain intensity at baseline and after 3 months, and a mean difference of 0.35 between the intervention and control groups.[13] Thus, the required number of patients per group is approximately 87, for a two-sided significance level of $\alpha = 0.05$ and a power of $1 - \beta = 0.8$.

Given the low rate of missing data (approximately 5% of baseline data), no imputation methods were applied, and complete case analyses were conducted. All analyses were performed in R (version 4.4.2).

The study was conducted across four rheumatology centers in Germany, approved by the Ethical Committee of the Ruhr-University Bochum, Germany (reference number 23-7831) and registered in the German Clinical Trials Register (DRKS00037509). Informed written consent was obtained from all patients. All data used in this research are already included in the article.

RESULTS

Clinical assessments and app usage. A total of 320 patients were screened, and 136 patients were consecutively enrolled in the study, with 73 patients randomized to IG and 63 to SOC between September 2023 and June 2024 (Table 1; Figure 1). The sample size could not be reached within the planned timeframe of 6 months. The extension of the recruitment period had to be discontinued after 4 months for organizational reasons. In line with the elements of a pragmatic design, not all patients randomized to the IG activated or regularly used the ACT app, and levels of engagement varied. Within the IG, 44 patients (60.3%) actively used the ACT app, and 29 (39.7%) were categorized as inactive app users. Of the inactive users, 16 (55.2%) had access to the app but chose not to use it, and 13 (44.8%) did not receive the app activation code from their health insurance provider, leaving them unable to engage with the intervention. At the 12-week follow-up, 19 IG patients (43.2%) completed all units in the ACT app, whereas 25 (56.8%)

Table 1. Demographical characteristics and baseline assessments stratified by group and user activity*

Characteristics	n	All app users (n = 73)	Active users (n = 44)	Inactive users (n = 29)	SOC patients (n = 63)
Age, mean ± SD, y	136	46.8 ± 13	47.3 ± 12	46.2 ± 12.4	50.9 ± 11.4
Gender female, n (%)	136	36 (49.3)	23 (52.3)	13 (44.8)	27 (42.9)
Education, university level, n (%)	123	18 (24.7)	10 (22.7)	8 (27.6)	11 (17.5)
Employment, n (%)	130	51 (68.8)	29 (65.9)	16 (59.3)	35 (55.5)
Time since axSpA symptoms, mean ± SD, y	129	18.2 ± 14.2	17.7 ± 13.8	19.0 ± 15.1	20.3 ± 13.5
Time since axSpA diagnosis, mean ± SD, y	132	13.3 ± 13	13.2 ± 13.7	13.6 ± 12.4	15.2 ± 12.7
BMI, mean ± SD, kg/m^2	116	28.9 ± 6.7	29.0 ± 6.3	28.6 ± 7.5	28.2 ± 6.7
CRP, mean ± SD, mg/dl	133	0.6 ± 1	0.4 ± 0.8	0.8 ± 1.2	0.6 ± 1.2
HLA-B 27, n (%)	120	46 (63.0)	25 (67.7)	21 (75.0)	44 (70.0)
Charlson comorbidity index (0–29), mean ± SD	133	1.1 ± 1.2	1.1 ± 1.1	0.9 ± 1.3	1.2 ± 1.3
NSAID index (0–100), mean ± SD	131	18.7 ± 29.2	18.2 ± 31.6	19.6 ± 25.6	19.1 ± 30.4
Patients on csDMARDs, n (%)	136	2 (2.7)	1 (2.3)	1 (3.4)	3 (4.8)
Patients on bDMARDs, n (%)	136	47 (64.4)	29 (65.9)	18 (64.3)	42 (66.6)
Patients on tsDMARDs, n (%)	136	5 (6.8)	4 (9.1)	1 (3.4)	3 (4.8)
Previous bDMARDs, mean ± SD, n	131	1.6 ± 1.4	1.7 ± 1.5	1.5 ± 1.6	1.4 ± 1.1
MPI, mean ± SD					
Pain-related life interference (0–6)	110	3.4 ± 1.1	3.3 ± 1.3	3.7 ± 0.6	3.6 ± 1.0
Pain severity (0–6)	110	3.4 ± 1.1	3.1 ± 1.2	3.8 ± 0.9	3.84 ± 1.2
Affective distress (0–6)	110	3.2 ± 1.2	3.1 ± 1.3	3.4 ± 1.1	3.2 ± 1.2
Social support (0–6)	110	3.4 ± 1.9	3.3 ± 2.0	3.5 ± 1.7	3.8 ± 1.5
Life control (0–6)	110	3.8 ± 1.2	4.0 ± 1.2	3.2 ± 1.1	3.5 ± 1
ASDAS (0–10), mean ± SD	129	2.7 ± 0.8	2.6 ± 0.8	2.9 ± 0.7	2.9 ± 0.8
BASDAI (0–10), mean ± SD	134	5.1 ± 1.8	5.0 ± 1.9	5.3 ± 1.6	5.4 ± 1.9
PhGA (0–10), mean ± SD	129	3 ± 1.5	2.7 ± 1.6	3.3 ± 1.3	2.9 ± 1.2
BASFI (0–10), mean ± SD	133	4.6 ± 2.4	4.5 ± 2.5	4.8 ± 2.1	4.9 ± 2.3
ASAS Health Index (0–17), mean ± SD	135	8.1 ± 3.5	7.9 ± 3.7	8.3 ± 3.3	8.0 ± 3.8
Pain on NRS (0–10), mean ± SD	134	5.5 ± 1.8	5.4 ± 1.9	5.7 ± 1.7	5.8 ± 1.9
PainDETECT >18, n (%)	135	13 (29.5)	13 (29.5)	6 (21.4)	22 (36.0)
HADS-D (0–21), mean ± SD	134	7.9 ± 4.3	7.6 ± 4.5	8.3 ± 3.9	9 ± 5.3
HADS-A (0–21), mean ± SD	134	8.5 ± 4.4	8.3 ± 4.6	9.0 ± 4.1	9 ± 4.3
HADS-D ≥8, n (%)	134	39 (53.4)	21 (47.7)	18 (62.1)	34 (55.7)
HADS-A ≥8, n (%)	134	43 (58.9)	24 (54.5)	19 (65.5)	30 (49.2)
FIQ (0–100), mean ± SD	127	49.4 ± 16.1	49.6 ± 15.6	49.2 ± 17.1	52.7 ± 21.3
SF-36 (PCS) (0–100), mean ± SD	132	33.5 ± 8.3	33.5 ± 9.6	33.5 ± 6.1	32.6 ± 7.4
SF-36 (MCS) (0–100), mean ± SD	132	41.7 ± 12.2	42.6 ± 12.2	40.4 ± 12.2	42.1 ± 13.1
mSQUASH total activity score, mean ± SD	132	9,483 ± 4,878	9,473 ± 5,156	9,500 ± 4,484	9,137 ± 5,675

* ASAS, Assessment of Spondyloarthritis International Society; ASDAS, Ankylosing Spondylitis Disease Activity Score; axSpA, axial spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; bDMARD, biologic disease-modifying antirheumatic drug; BMI, body mass index; CRP, C-reactive protein; csDMARD, conventional synthetic DMARD; FIQ, Fibromyalgia Impact Questionnaire; HADS-A, Hospital Anxiety and Depression Scale – anxiety subscale; HADS-D, HADS – depression subscale; MCS, mental component summary; MPI, Multidimensional Pain Inventory; mSQUASH, modified Short Questionnaire to Assess Health-enhancing physical activity; NRS, numerical rating scale; NSAID, nonsteroidal anti-inflammatory drug; PCS, physical component summary; PhGA, Physician Global Assessment; SF-36, 36-item Short Form; SOC, standard of care; tsDMARD, targeted-synthetic DMARD.

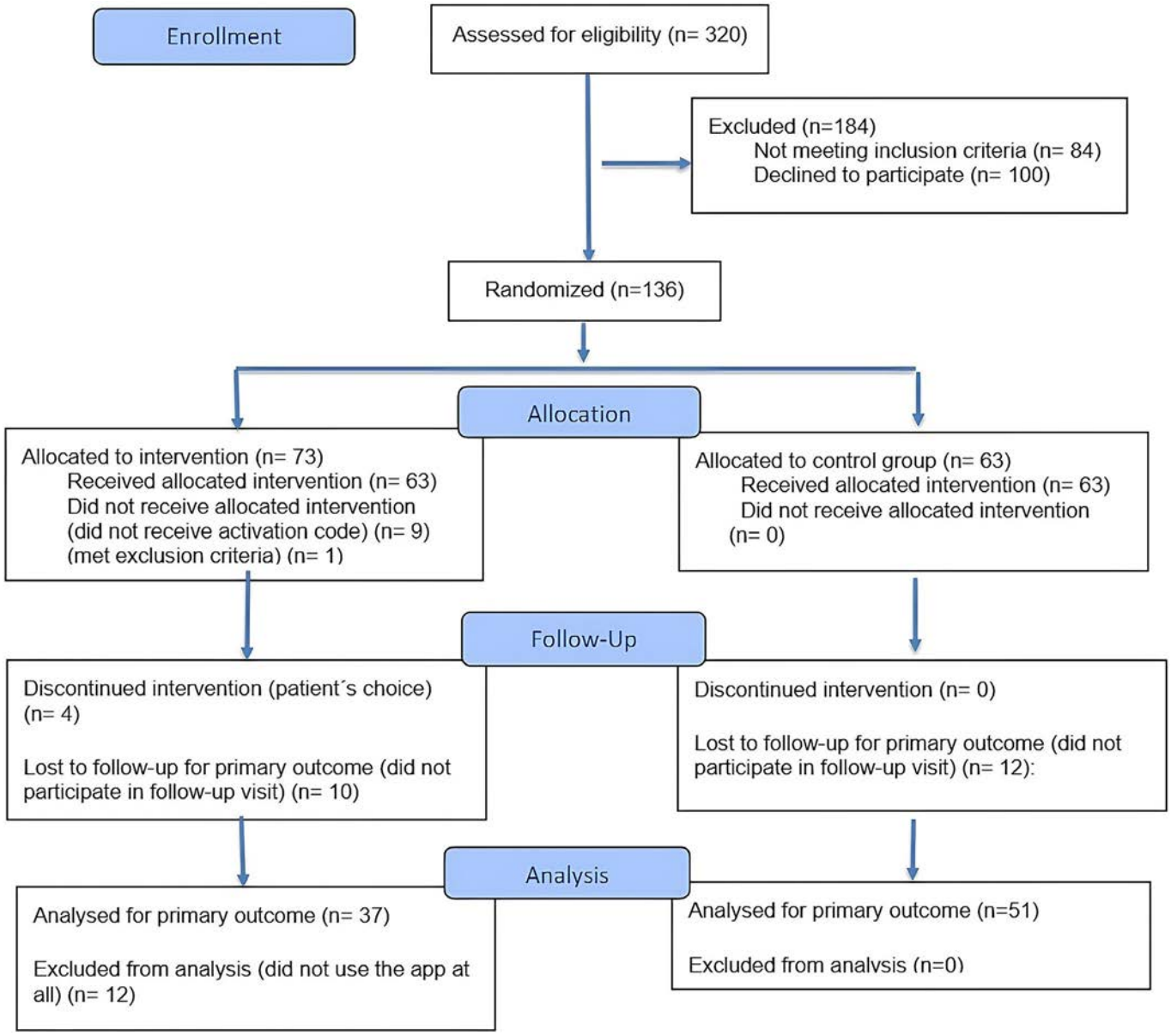


Figure 1. Study flowchart.

used the ACT app actively but did not complete all seven units. Based on electronic records, active app users completed an average of 70% of the units. (Table 1).

Baseline characteristics, including demographic and clinical data, were comparable between the IG and SOC group. The mean ± SD age was 46.8 ± 13.0 years in the IG and 50.9 ± 11.4 years in the SOC group. Gender distribution was similar, with 49.3% of the IG and 42.9% of the SOC group being female. Time since first symptoms and time since diagnoses were comparable. Importantly, there was no difference in the usage of NSAIDs and b/tsDMARDs or the comorbidity index (Table 1). The CRP level was also comparable.

At baseline, IG and SOC showed similar baseline values of the MPI dimensions. In detail, the mean ± SD results for pain-related life interference (3.4 ± 1.1 vs 3.6 ± 1.0), pain severity (3.4 ± 1.1 vs 3.8 ± 1.2), affective distress (3.2 ± 1.2 in both groups), social support (3.4 ± 1.9 vs 3.8 ± 1.5), and life control (3.8 ± 1.2 vs 3.5 ± 1.0) were comparable between the IG and SOC group (Table 1). Furthermore, baseline measures for disease activity (ASDAS, BASDAI, and PhGA), physical and global functioning (BASFI, PCS, MCS, and ASAS-HI), spinal mobility (Bath Ankylosing Spondylitis Metrology Index) and measurements of pain (pain on NRS and Pain DETECT), and physical activity (mSQUASH) as well as Patient Acceptable Symptom State, HADS, FIQ, and Patient-Activation Measure-13 results showed no relevant differences between the IG and SOC group (Table 1).

Improvements in pain-related life interference (–0.36; 95% confidence interval [CI] –0.73 to 0.01) and affective distress related to

Table 2. Effect of the digital health application on primary and secondary outcomes (analysis of covariance analyses adjusted for outcome at baseline)*

Outcome	β	95% CI
Pain-related interference, 0–6	−0.36	−0.73 to 0.01
Emotional Distress, 0–6	−0.40	−0.84 to 0.03
ASDAS, 0–10	−0.09	−0.37 to 0.20
BASDAI, 0–10	−0.76	−1.3 to −0.25
PhGA, 0–10	−0.56	−0.98 to −0.14
BASFI, 0–10	−0.65	−1.1 to −0.19
Pain, NRS 0–10	−0.61	−1.3 to 0.10
PainDETECT, 0–38	−2.13	−4.1 to −0.12
Pain severity (MPI), 0–6	−0.12	−0.48 to 0.25
Social support (MPI), 0–6	0.07	−0.43 to 0.57
Life control (MPI), 0–6	0.05	−0.45 to 0.54
HADS-D, 0–21	−0.92	−1.9 to 0.07
HADS-A, 0–21	−0.77	−1.9 to 0.33
FIQ, 0–100	−4.66	−10 to 0.65
ASAS Health Index, 0–17	−1.32	−2.3 to −0.35
SF-36 (PCS), 0–100	2.15	−0.32 to 4.6
SF-36 (MCS), 0–100	1.74	−1.5 to 5.0
mSQUASH, total activity score	−475.47	−2,016 to 1,065

* ASAS, Assessment of Spondyloarthritis International Society; ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; CI, confidence interval; FIQ, Fibromyalgia Impact Questionnaire; HADS-A, Hospital Anxiety and Depression Scale – anxiety subscale; HADS-D, HADS – depression subscale; MCS, mental component summary; MPI, Multidimensional Pain Inventory; mSQUASH, modified Short Questionnaire to Assess Health-enhancing physical activity; NRS, numerical rating scale; PCS, physical component summary; PhGA, Physician Global Assessment; SF-36, 36-item Short Form.

the disease (–0.4; 95% CI –0.84 to 0.03) were greater in IG compared with SOC (Table 2). Improvements in most secondary disease-specific measures were more pronounced in active app users compared with SOC (Table 2; Figure 2). In particular, BASDAI (–0.76; 95% CI –1.3 to –0.25), PhGA (–0.56; 95% CI –0.98 to –0.14), BASFI (–0.65; 95% CI –1.1 to –0.19), PainDETECT (–2.13; 95% CI –4.1 to –0.12), and ASAS-HI (–1.32; 95% CI –2.3 to –0.35) showed a greater improvement in IG compared with SOC. The percentage of patients with clinically relevant HADS – depression subscale scores (≥8) reduced from 46.5% to 32.4% in IG and from 55.7% to 46.8% in SOC, whereas the percentage of patients with clinically relevant anxiety was only reduced in IG (53.5% to 43.2%) and slightly increased in SOC (49.2% to 53.2%).

A higher proportion of active app users had a subjective perception of improvement (PGIC) over 12 weeks in contrast to SOC patients (36 vs 13%). Conversely, more participants in the SOC group had a subjective perception of worsening compared with active app users (35 vs 11%) (Figure 3).

App evaluation metrics. The DHA was promoted by five patients (14.7%), whereas the others were either passive users (n = 9, 26.5%) or detractors (n = 20, 58.9%). The quality of the ACT app was evaluated as being good with a mean ± SD MARS score of 3.8 ± 0.5 (Figure 4A). The best mean ± SD MARS rating was achieved in the functionality domain (4.3 ± 0.7) followed by the information domain (4.1 ± 0.6). In the subjective quality domain, the app achieved the weakest rating, with a mean ± SD score of 2.9 ± 0.9. The effect of the app on patients' health behavior was rated as moderate with a mean ± SD score of 3.3 ± 1.0. The usability was rated favorable with a mean ± SD score of 73.3 ± 17.0.

We identified minor barriers to the use of the app (Figure 4B). Moderate challenges were related to *implementation into therapy* (3.7 ± 3.1), *user adherence* (3.6 ± 2.7), and the *app not meeting individual needs* (3.1 ± 2.7). Minor barriers included *required digital expertise* (2.2 ± 2.7), *impact on the doctor-patient relationship* (2.1 ± 2.4), usability (1.8 ± 2.2), *validity* in meeting user expectations (1.4 ± 1.8), *technical problems to access* (1.2 ± 0.8), *data protection concerns* (1.2 ± 2.3) and *costs* (0.7 ± 1.6) (Figure 4B).

DISCUSSION

Our study investigated the effectiveness of the HelloBetter app, a DHA, in addressing multidimensional pain in patients with axSpA and persistent pain. The primary outcome of the study, the reduction in pain-related life interference measured by the MPI, indicated a greater improvement in IG compared with SOC. Although the absolute change in pain-related life interference was small, it may still be clinically relevant in a population with stable disease control under optimized pharmacologic treatment, as even modest improvements can have a meaningful impact when achieved through a low-risk and easily accessible intervention such as a DHA.

A recent meta-analysis found that ACT reduced pain intensity after treatment compared with applied relaxation, with a standardized mean difference of –0.26 (95% CI –0.58 to 0.07), consistent with the small change observed in our study.[12] This is consistent with earlier studies highlighting the efficacy of DHAs in addressing chronic back pain through psychological and behavioral strategies,[26] whereas there was no effect on neuropathic pain in our study. The observed reduction in emotional distress and improvements in disease-specific QoL measures and the slight improvement in the PGIC further supports the utility of DHAs in managing complex pain syndromes in axSpA, including nociplastic pain.[27] ACT has the potential to enhance psychological flexibility and improve patient outcomes by focusing on helping individuals accept their pain and commit to actions that align with their values, thereby reducing the psychological burden associated with chronic conditions.[28]

Studies indicate that ACT can lead to improve several outcomes including physical functioning in patients with axSpA. A recent study demonstrated that ACT-based digital interventions effectively improve physical functioning.[27] Moreover, integrating ACT into DHAs offers a new approach to delivering psychological interventions. Digital platforms can provide accessible, cost-

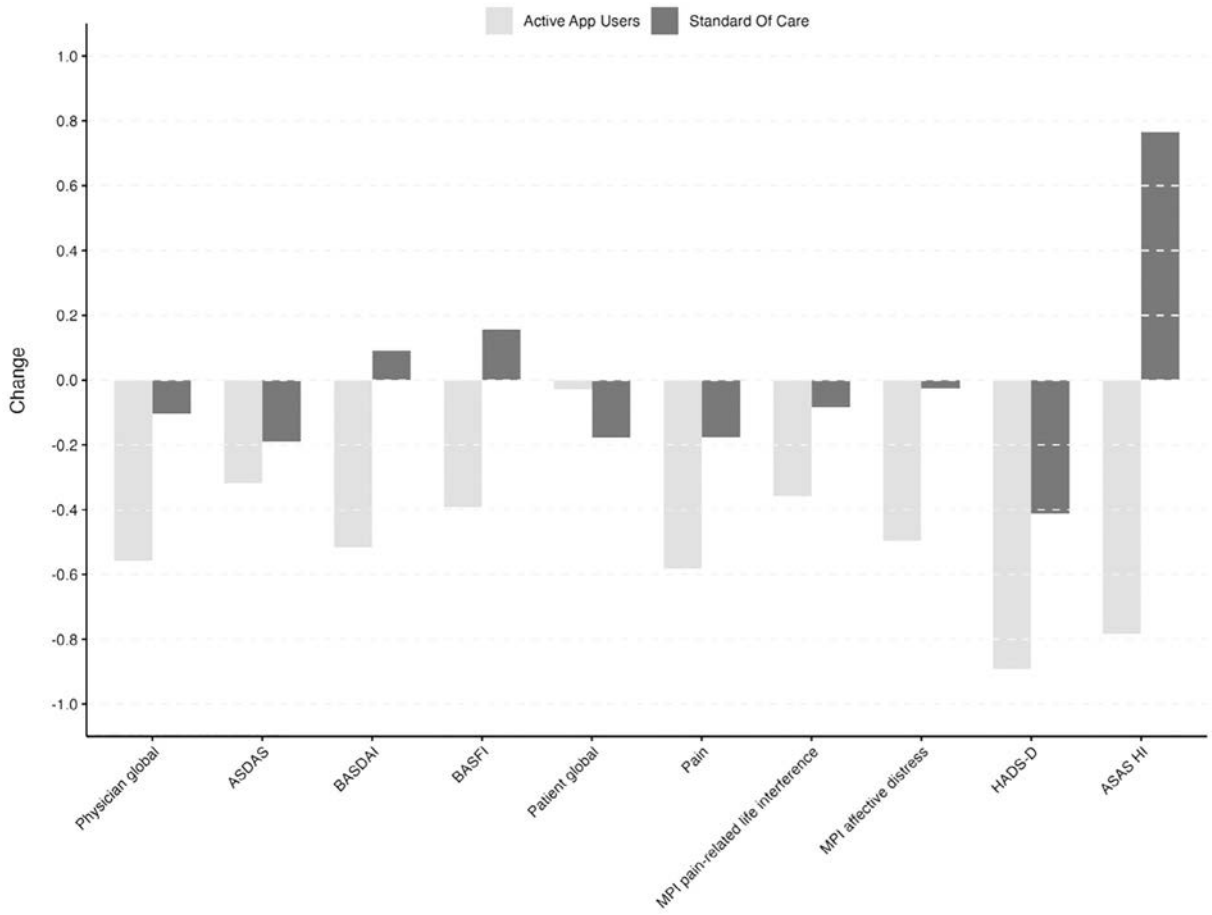


Figure 2. Change of clinical variables stratified by group. Mean change of assessments compared between active app users and standard of care patients. ASAS-HI, Assessment of Spondyloarthritis International Society Health Index; ASDAS, Ankylosing Spondylitis Disease Activity Score; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; HADS-D, Hospital Anxiety and Depression Scale – depression subscale; MPI, Multidimensional Pain Inventory.

effective, and user-friendly means for patients to engage with ACT principles, potentially enhancing adherence and engagement.[29] This is particularly relevant in rheumatology, in which patients may face barriers to accessing traditional in-person therapy sessions.[30] The incorporation of ACT and digital health into comprehensive treatment plans for axSpA aligns with current recommendations emphasizing the importance of addressing both the physical and psychological aspects.[31,32] ACT empowers patients to manage their symptoms more effectively, leading to improved overall well-being.[13,17] The minimal improvement observed in ASDAS and BASDAI was marginal and did not reach the threshold for minimal clinical importance. Therefore, we cannot conclude that ACT has an effect on disease activity.[33,34]

Our findings also align with evidence from studies on other digital health solutions, which demonstrated significant reductions in back pain and improved functional outcomes among patients with chronic conditions.[26,27] Another recent study demonstrated the effectiveness of a digital ACT program in reducing symptom burden and improving PROs in fibromyalgia, a condition often overlapping with axSpA.[35–38] Previous evidence supported the use of DHAs in persistent pain management through psychological and behavioral strategies. A recent study highlighted the increasing acceptance and use of digital health technologies among rheumatology patients, emphasizing their utility in enhancing patient-centered care and improving access to therapeutic interventions.[39]

The SOC group, which followed traditional rheumatologist-guided care without the use of the app, demonstrated no significant improvements in MPI dimensions. This highlights the limitations of standard care in addressing the multidimensional aspects of persistent pain. By integrating behavioral therapy into its framework, the HelloBetter app provides a more comprehensive approach to pain management, addressing also the psychological components of persistent pain. Other DHAs have similarly shown their utility in bridging the gaps left by conventional care.

The results of this study highlight the transformative potential of DHAs in the clinical management of axSpA. The HelloBetter app's ACT-based framework aligns with current recommendations for

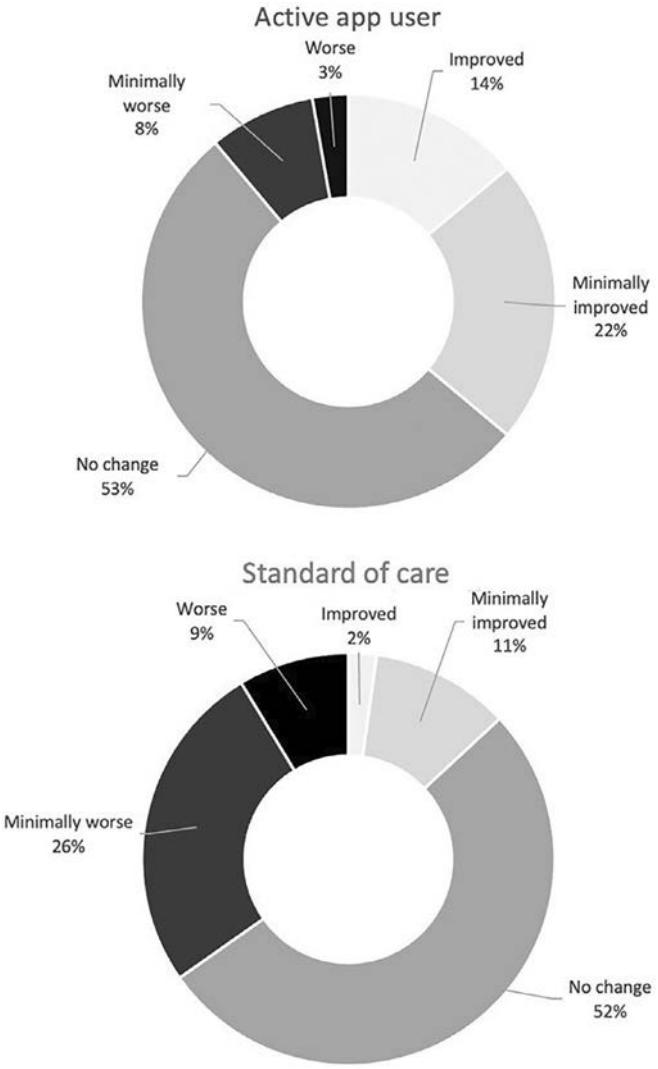


Figure 3. Patient Global Impression of Change categories: comparison between active app users and the standard of care group. All categories are given as percentages. The two categories at the end of the 7-point Likert scale ("very much improved"/"much improved" and "very much worse"/"much worse") have been grouped together resulting in five categories ranging from "improved" to "worse."

managing RMDs and also recommendations for the management of fibromyalgia and related disorders, which emphasize the importance of psychological and behavioral interventions.[9,40] By enabling patients to engage with evidence-based therapies from the comfort of their homes, DHAs remove many of the logistical barriers associated with traditional care models, such as transportation difficulties and time constraints, which are known barriers for patients.[30] For clinicians, the integration of DHAs offers an opportunity to enhance patient self-management. Given the limited time available in routine clinical practice to deliver comprehensive ACT, DHAs can help save valuable time during consultations while still enabling effective implementation of ACT principles and improving pain management outcome.[39,41,42]

In addition, the 2021 EULAR recommendations for the implementation of self-management strategies in patients with inflammatory arthritis and the 2024 EULAR points to consider for patient education in physical activity and self-management of pain during transitional care state that digital health care can help patients to self-manage and should be considered for inclusion in supported self-management where appropriate and available. They also state that patient education during transitional care should consist of a variety of learning formats, including digital health, and that digital solutions seem to be both feasible and beneficial for patients with RMDs, especially for younger patients.[31,32,43,44]

We have to acknowledge that the DHA did not provide a clear benefit for all patients, as perceived improvements were minimal with a moderate perceived positive influence of the DHA on user's life, although technical functionality information quality and usability were rated positively. Adherence to digital interventions remains a critical challenge, with only 43% of participants in the IG completing all lessons in the app. This adherence rate mirrors findings from similar studies involving digital therapeutics, highlighting the need for enhanced user engagement strategies.[45] Barriers to adherence often include low digital literacy, lack of motivation, and insufficient technical support. In contrast, patients state that they are willing to use digital applications, such as DHAs.[42,46–48] Furthermore, onboarding processes should be optimized, including detailed user education and ongoing technical support, to ensure patients feel confident using the app.[42,48]

In line with this, several challenges were encountered during our study, including the relatively low activation rate of the app within the first 4 weeks, which necessitated follow-up interventions to assist with onboarding. Moreover, excluding patients who could not access the DHA because of insurer delays may underestimate real-world barriers.

Furthermore, even when the cost of the HelloBetter app was covered by insurance companies, some patients did not receive an activation code from their insurers. This could represent a potential barrier to the successful implementation and widespread adoption of DHAs in the future. Another limitation of our study is that we were unable to reach the sample size of 87 patients per group. This is based on the high screening failure rate of 42% with a substantial number of patients unwilling to use an app. Another limitation is the unblinded design of the study, which may have influenced both PROs and the interpretation of results. Future studies should aim to incorporate physician-blinded study designs, as patient blinding is not feasible with DHAs. In addition, the open-label design may also have introduced expectancy bias in PROs.

The generalizability of our findings to settings outside Germany is limited, as prescription-based, health insurance–funded DHAs currently exist only in Germany. Nevertheless, other types of digital tools to support patients are available internationally. Unlike many freely available or commercial apps, DHAs are

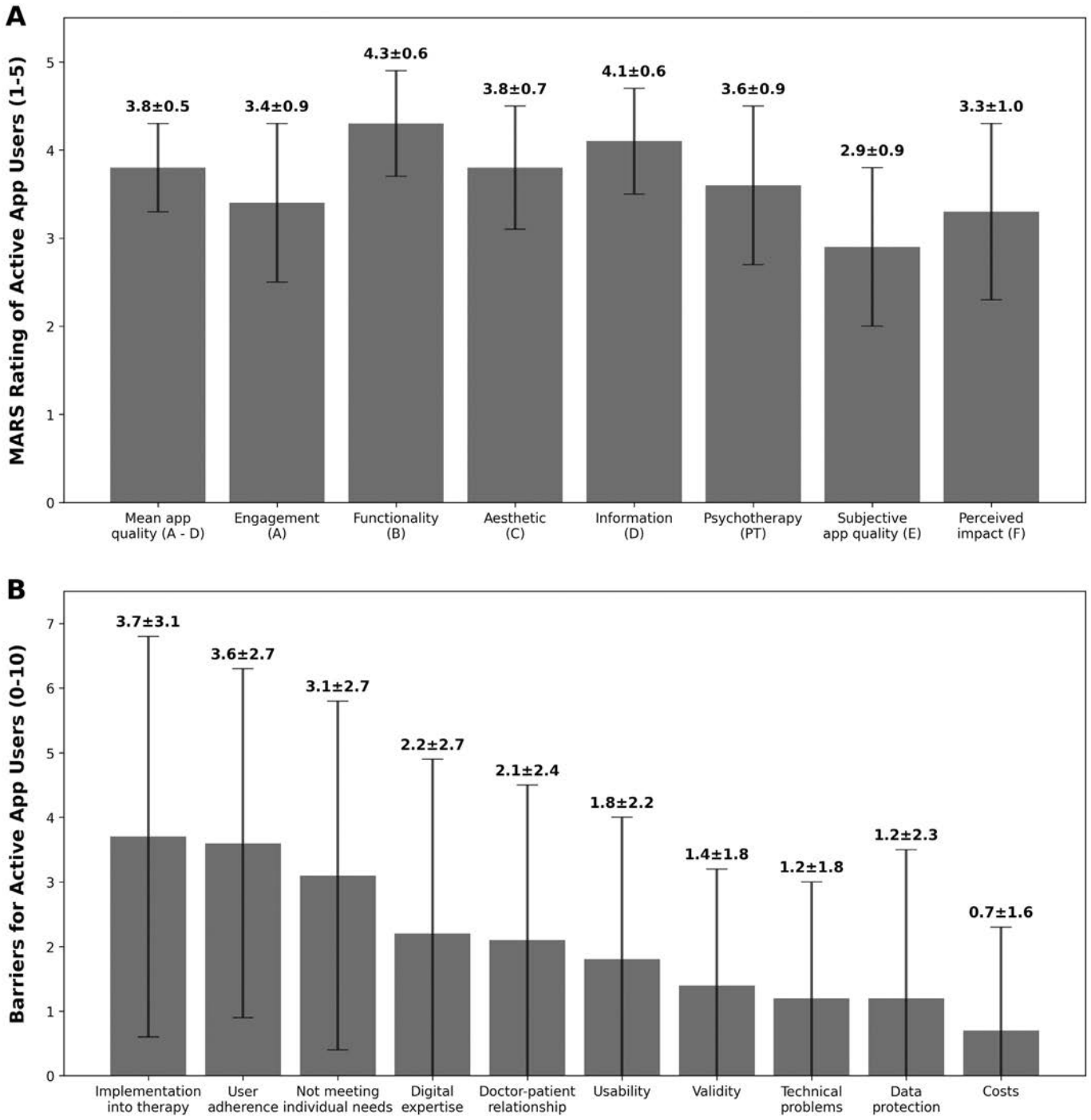


Figure 4. Active app user's (A) MARS rating and (B) barriers to using the app. (A) Mean ± SD values of the overall MARS rating of active app users. (B) Mean ± SD values of the barriers questionnaire (score 1–10) of active app users ranged from the largest to smallest barrier. MARS, mobile app rating scale.

subject to governmental approval by the BfArM and must demonstrate efficacy and safety according to defined regulatory requirements. The underlying therapeutic concept and delivery format, however, may be transferable to other health care systems, particularly where there is increasing interest in integrating evidence-based digital interventions into routine care. Future research should focus on optimizing the design and functionality of DHAs to enhance user engagement and long-term adherence and may include gamification.[49] Economic evaluations are also needed to assess the cost-effectiveness of DHAs in pain management. Additionally, studies should investigate the long-term impact of DHAs on disease progression, comorbidities, and overall QoL.

The digital intervention demonstrated reductions in pain-related life interference and emotional distress in patients with axSpA, supporting its potential as a supplementary tool in pain management. These findings contribute to the growing body of evidence supporting the use of DHAs in rheumatology and underscore their role in addressing the multidimensional nature of persistent pain. By fostering greater accessibility, personalized care, and patient empowerment, DHAs represent a transformative advancement in the management of complex chronic conditions.

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017; 390(10089):73–84.
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(1):19–34.
3. Kieskamp SC, Paap D, Carbo MJG, et al. Central sensitization has major impact on quality of life in patients with axial spondyloarthritis. Semin Arthritis Rheum 2022;52:151933.
4. Lee YC, Malfait AM, Ogdie AR. Unmet needs in spondyloarthritis: understanding and managing chronic pain. J Rheumatol 2024; 51(12):1235–1240.
5. Kiltz U, Hoeper K, Hammel L, et al. Work participation in patients with axial spondyloarthritis: high prevalence of negative workplace experiences and long-term work impairment. RMD Open 2023;9(1): e002663.
6. Poddubnyy D, Callhoff J, Spiller I, et al. Diagnostic accuracy of inflammatory back pain for axial spondyloarthritis in rheumatological care. RMD Open 2018;4(2):e000825.
7. Drouet J, Gossec L, Jacquemin C, et al. Fluctuation of pain is frequent in rheumatoid arthritis and axial spondyloarthritis: a 12 weeks prospective study of 165 patients. Joint Bone Spine 2022;89(3):105306.
8. Lumley MA, Schubiner H, Lockhart NA, et al. Emotional awareness and expression therapy, cognitive behavioral therapy, and education for fibromyalgia: a cluster-randomized controlled trial. Pain 2017; 158(12):2354–2363.
9. Macfarlane GJ, Kronisch C, Atzeni F, et al. EULAR recommendations for management of fibromyalgia. Ann Rheum Dis 2017;76(12):e54.
10. Catella S, Gendreau M, Vega N, et al. Clinical impact of a digital behavioral therapy for fibromyalgia management: a randomized controlled trial. Abstract presented at: ACR Convergence 2021; November 6, 2021. Accessed March 15, 2025. https://acrabstracts.org/abstract/clinical-impact-of-a-digital-behavioral-therapy-for-fibromyalgia-management-a-randomized-controlled-trial/
11. Sandal LF, Bach K, Øverås CK, et al. Effectiveness of app-delivered, tailored self-management support for adults with lower back pain-related disability: a selfBACK randomized clinical trial. JAMA Intern Med 2021;181(10):1288–1296.
12. Hughes LS, Clark J, Colclough JA, et al. Acceptance and commitment therapy (ACT) for chronic pain: a systematic review and meta-analyses. Clin J Pain 2017;33(6):552–568.
13. Lin J, Faust B, Ebert DD, et al. A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018;20(8):e244.
14. de Thurah A, Bosch P, Marques A, et al. 2022 EULAR points to consider for remote care in rheumatic and musculoskeletal diseases. Ann Rheum Dis 2022;81(8):1065–1071.
15. Bundesinstitut für Arzneimittel und Medizinprodukte. HelloBetter Chronic Pain. Accessed January 4, 2025. https://diga.bfarm.de/de/verzeichnis/01304
16. Lin J, Paganini S, Sander L, et al. An internet-based intervention for chronic pain. Dtsch Arztebl Int 2017;114(41):681–688.
17. Baumeister H, Seifferth H, Lin J, et al. Impact of an acceptance facilitating intervention on patients' acceptance of internet-based pain interventions: a randomized controlled trial. Clin J Pain 2015;31(6): 528–535.
18. Sander L, Paganini S, Lin J, et al. Effectiveness and cost-effectiveness of a guided internet- and mobile-based intervention for the indicated prevention of major depression in patients with chronic back pain-study protocol of the PROD-BP multicenter pragmatic RCT. BMC Psychiatry 2017;17(1):36.
19. Paganini S, Lin J, Kählke F, et al. A guided and unguided internet- and mobile-based intervention for chronic pain: health economic evaluation alongside a randomised controlled trial. BMJ Open 2019;9(4): e023390.
20. Rifbjerg-Madsen S, Wæhrens EE, Danneskiold-Samsøe B, et al. Psychometric properties of the painDETECT questionnaire in rheumatoid arthritis, psoriatic arthritis and spondyloarthritis: rasch analysis and test-retest reliability. Health Qual Life Outcomes 2017;15(1):110.
21. Flor H, Rudy TE, Birbaumer N, et al. The applicability of the West Haven-Yale Multidimensional Pain Inventory in German-speaking countries. Data on the reliability and validity of the MPI-D. Article in German. Schmerz 1990;4(2):82–87.
22. Reyes A, Qin P, Brown CA. A standardized review of smartphone applications to promote balance for older adults. Disabil Rehabil 2018;40(6):690–696.
23. Adams C, Walpola R, Schembri AM, et al. The ultimate question? Evaluating the use of Net Promoter Score in healthcare: a systematic review. Health Expect 2022;25(5):2328–2339.
24. Dworkin RH, Turk DC, Farrar JT, et al; IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113(1-2):9–19.
25. Bangor A, Kortum P, Miller J. An empirical evaluation of the system usability score. Int J Human-Comp Int 2008;24(6):574–594.
26. Boy K, May S, Labinsky H, et al. Back on track - digital health applications to treat back pain of rheumatic patients? Results of a qualitative interview study. Rheumatol Int 2024;44(12):2977–2987.
27. Blaskowitz PPVA, Liphardt AM, Bouzas C, et al. Impact of the digital health application ViViRA on spinal mobility, physical function, quality of life and pain perception in spondyloarthritides patients: a randomized controlled trial. Arthritis Res Ther 2024;26(1):208.
28. McCracken LM. Psychological flexibility, chronic pain, and health. Annu Rev Psychol 2024;75(1):601–624.

29. Kristjánsdóttir OB, Fors EA, Eide E, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013;15(1):e5.
30. Fongen C, Sveaas SH, Dagfinrud H. Barriers and facilitators for being physically active in patients with ankylosing spondylitis: a cross-sectional comparative study. Musculoskeletal Care 2015;13(2):76–83.
31. 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(10):1278–1285.
32. Courel-Ibáñez J, Prieto-Moreno R, Briones-Vozmediano E, et al. EULAR points to consider for patient education in physical activity and self-management of pain during transitional care. Ann Rheum Dis Published online November 11, 2024. https://doi.org/10.1136/ard-2024-226448
33. Zochling J. Measures of symptoms and disease status in ankylosing spondylitis: Ankylosing Spondylitis Disease Activity Score (ASDAS), Ankylosing Spondylitis Quality of Life Scale (ASQoL), Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), Bath Ankylosing Spondylitis Functional Index (BASFI), Bath Ankylosing Spondylitis Global Score (BAS-G), Bath Ankylosing Spondylitis Metrology Index (BASMI), Dougados Functional Index (DFI), and Health Assessment Questionnaire for the Spondylarthropathies (HAQ-S). Arthritis Care Res (Hoboken) 2011;63(S11)(suppl 11):S47–S58.
34. Machado P, Landewé R, Lie E, et al; Assessment of SpondyloArthritis international Society. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. Ann Rheum Dis 2011;70(1):47–53.
35. Gendreau RM, McCracken LM, Williams DA, et al. Self-guided digital behavioural therapy versus active control for fibromyalgia (PROSPER-FM): a phase 3, multicentre, randomised controlled trial. Lancet 2024;404(10450):364–374.
36. Wach J, Letroublon MC, Coury F, et al. Fibromyalgia in spondyloarthritis: effect on disease activity assessment in clinical practice. J Rheumatol 2016;43(11):2056–2063.
37. Gok K, Cengiz G, Erol K, et al. Neuropathic pain component in axial spondyloarthritis and the influence on disease burden. J Clin Rheumatol 2018;24(6):324–327.
38. Redeker I, Callhoff J, Hoffmann F, et al. The prevalence and impact of comorbidities on patients with axial spondyloarthritis: results from a nationwide population-based study. Arthritis Res Ther 2020;22(1):210.
39. May S, Darkow R, Knitza J, et al. Digital transformation of rheumatology care in Germany: cross-sectional national survey. J Med Internet Res 2025;27:e52601.
40. Kundakci B, Hall M, Atzeni F, et al. International, multidisciplinary Delphi consensus recommendations on non-pharmacological interventions for fibromyalgia. Semin Arthritis Rheum 2022;57:152101.
41. Knitza J, Tascilar K, Messner EM, et al. German mobile apps in rheumatology: review and analysis using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2019;7(8):e14991.
42. Knitza J, Simon D, Lambrecht A, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020;8(8):e19661.
43. 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 (Oxford) 2021;60(1):196–206.
44. 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(1):48.
45. Kempin R, Richter JG, Schlegel A, et al. Monitoring of disease activity with a smartphone app in routine clinical care in patients with axial spondyloarthritis. J Rheumatol 2022;49(8):878–884.
46. Kiefer D, Braun J, Kiltz U, et al. Patients' awareness towards physical activity in the treatment of axial spondyloarthritis. Joint Bone Spine 2023;90(5):105585.
47. Strunz PP, Maire ML, Heusinger T, et al. Apps in rheumatology: is there a need for an app in therapy for axial spondyloarthritis? Article in German. Z Rheumatol 2023;82(3):256–261.
48. Labinsky H, Gupta L, Raimondo MG, et al. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023;43(4):713–719.
49. 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(5):314–331.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 96–105
DOI 10.1002/acr.25675

Impact of Social Factors on Visual Analog Scale Pain in Patients With Spondyloarthritis: A Propensity Score Matching Analysis

Luis Ángel Calvo Pascual,[1] David Castro Corredor,[2] Eduardo Collantes-Estévez,[3] and Clementina López-Medina[3]

Objective. Pain in spondyloarthritis reflects not only inflammatory activity but also social context. We quantified the association between a composite Social Factors Index (SFI) and patient-reported pain severity on the Visual Analog Scale (VAS).

Methods. We analyzed 2,042 patients from REGISPONSER (N = 1,514) and RESPONDIA (N = 528). Missing data were handled with multiple imputation (m = 20). The SFI combined education, employment, housing, income source, and Graffar class using penalized regression (sparse-group lasso). Mutual information (MI) quantified the SFI's relative importance. Propensity scores (logistic regression) supported 1:1 caliper matching (0.2 SD on the logit), comparing the highest (Q4) and lowest (Q1) SFI quartiles.

Results. Before matching, Q1 and Q4 included on average 918 and 829 patients per imputation, differing mainly in race (White 80.8% vs 86.8%; standardized mean difference [SMD] = −0.16) and country (Spain 66.2% vs 78.2%; SMD = −0.27). After matching, a mean of 745 pairs (range 731–761) were retained with excellent covariate balance (all $|SMD| < 0.10$). In the matched sample, patients in Q4 reported higher pain (average treatment effect on the treated +0.38 VAS points; 95% confidence interval [CI] 0.05–0.71; $P = 0.023$). Postmatching regression yielded an identical estimate ($\beta = +0.37$; 95% CI 0.04–0.70; $P = 0.030$). MI indicated the SFI carried more information about VAS than any single social variable. Stratified analyses suggested heterogeneity: Spain +0.31 (95% CI 0.08–0.71) and Latin America +0.62 (95% CI −0.22 to 1.45).

Conclusion. Social disadvantage, summarized by the SFI, is independently associated with greater pain severity in spondyloarthritis after adjustment. These findings support incorporating social determinants into clinical assessment and pain management.

INTRODUCTION

Spondyloarthritis (SpA) encompasses a group of chronic inflammatory rheumatic diseases that primarily affect the axial skeleton, leading to significant physical limitations and diminished quality of life. It often strikes individuals in early adulthood, causing chronic back pain, stiffness, and progressive loss of mobility. Consequently, SpA is associated with severe physical limitations, functional impairment, and decreased quality of life.[1]

Pain is a central and enduring symptom of SpA, traditionally attributed to inflammatory processes. However, a substantial subset of patients continues to experience significant pain despite effective control of inflammation through pharmacologic interventions. This persistent pain suggests the involvement of noninflammatory mechanisms, notably central sensitization, an increased responsiveness of nociceptive neurons in the central nervous system, leading to heightened pain sensitivity and perception.[2]

Beyond biologic mechanisms, social determinants have emerged as important contributors to pain severity in SpA.[3,4] Lower educational attainment and socioeconomic disadvantage are independently associated with higher patient-reported pain, even after adjustment for disease activity and treatment access.[5–7] In a cluster analysis, patients in the low socioeconomic group experienced longer diagnostic delays, higher body mass

[1]Universidad Pontificia Comillas, Madrid, Spain; [2]Hospital General Universitario de Ciudad Real, Ciudad Real, Spain; [3]Maimonides Institute for Research in Biomedicine of Cordoba, Reina Sofia University Hospital, University of Cordoba, Córdoba, Spain.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25675).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25675.

Address correspondence via email to Luis Ángel Calvo Pascual, BSc, MSc, PhD, at lacalvo@comillas.edu.

Submitted for publication May 1, 2025; accepted in revised form October 2, 2025.

SIGNIFICANCE & INNOVATIONS

- We developed a Social Factors Index (SFI) using sparse-group lasso with bootstrap across multiply imputed datasets; mutual information quantified the SFI's informativeness for patient-reported Visual Analog Scale (VAS) pain.
- The SFI captured more mutual information with VAS than any single social variable, including Graffar Scale class, employment status, housing conditions, or income.
- Using propensity score matching (>700 matched pairs per imputation), patients in the highest SFI quartile reported +0.38 VAS points versus the lowest (95% confidence interval 0.05–0.71; $P = 0.023$), supporting an association independent of measured disease activity.
- Findings underscore the clinical relevance of social vulnerability in spondyloarthritis and support integrating social determinants into comprehensive care and management.

index, and more severe structural damage despite adjustment for disease duration; although access to biologic disease-modifying antirheumatic drugs (DMARDs) was comparable across groups, disadvantaged patients had a substantially higher prevalence of permanent work disability, defined as official certification of long-term incapacity for work, independent of disease activity and duration.[8] Moreover, patients with chronic low back pain reported significantly worse pain severity, disability, and health-related quality of life than those with axial SpA, despite similar pain intensity scores, underscoring that pain experience extends beyond inflammatory burden.[9]

The present study investigates whether social determinants contribute to patient-reported pain in SpA using data from two registries (REGISPONSER and RESPONDIA). We examined education, employment status, housing conditions, and Graffar social class and aggregated them into a Social Factors Index (SFI) whose component weights were learned via lasso regularization (sparse-group lasso) with bootstrap resampling across multiply imputed datasets. We then quantified the importance of the SFI for Visual Analog Scale (VAS) pain using mutual information (MI) and applied propensity score matching to balance key demographic and clinical covariates, contrasting patients in the highest and lowest SFI quartiles.

PATIENTS AND METHODS

Study design. This study was based on data obtained from two multicenter observational registries: REGISPONSER and RESPONDIA, both specifically designed to collect standardized clinical information on patients with SpA.

REGISPONSER is a national, multicenter registry from Spain that prospectively enrolled consecutive patients diagnosed with SpA according to the European Spondyloarthropathy Study Group (ESSG) criteria.[10] Recruitment was conducted between March 2004 and March 2007 across 31 rheumatology departments. Detailed information regarding the design, sampling methodology, and recruitment process can be found in Collantes et al.[11]

RESPONDIA, a parallel registry with an identical study design, was conducted between 2006 and 2007 in Latin America. It included 33 centers across eight countries: Argentina, Chile, Colombia, Costa Rica, Mexico, Uruguay, and Venezuela. Like REGISPONSER, RESPONDIA included consecutive patients who met the ESSG classification criteria for SpA and used the same case report forms, clinical definitions, and variable collection procedures.[12] For this study, data from both sources were merged to generate a large, multinational cohort of patients with SpA.

Inclusion and exclusion criteria. All patients aged ≥18 years with a diagnosis of SpA according to the ESSG criteria were eligible. Participants were drawn from the REGISPONSER registry (Spain, 2004–2007) and the RESPONDIA registry (Latin America, 2006–2007). Patients with incomplete clinical information in key exposure or outcome variables, or those with duplicate records, were excluded. No additional exclusion criteria were applied.

Outcomes. The primary outcome of interest was pain intensity, measured using the VAS, which ranged from 0 (no pain) to 10 (worst imaginable pain), referring to overall pain experienced during the previous week. No formal secondary outcomes were prespecified.

Variables. This study analyzed data from 2,042 patients diagnosed with SpA. A total of 66 variables were collected and organized into four conceptual domains: personal variables, social variables, treatment-related variables, and disease activity. All variables were assessed at the baseline visit of each registry, with laboratory tests performed within 15 days before inclusion. Detailed information is provided in the Supplementary File 1, which includes six tables, three with categorical variables and three with numerical variables, each presented for the full sample, for REGISPONSER, and for RESPONDIA.

The personal variables include age, weight, height, physical activity habits, smoking status, race and ethnicity, and country of residence across Ibero-American regions such as Spain, Argentina, and Mexico. Genetic and familial factors were captured through HLA-B27 status (positive or negative) and the presence or absence of a family history of SpA.

The social variables domain was composed of education attainment, ranging from illiterate to university level. Employment status included categories such as employed, unemployed, student, retired, or homemaker. Socioeconomic stratification was

measured using the Graffar scale,[13] and housing conditions ranged from inadequate to excellent, with or without luxury. Marital status, sources of income (eg, salary, profits, donations, or rents), and professional background (eg, skilled or unskilled labor, technician, or employee) were also recorded. All social variables were systematically collected for all participants at the baseline visit.

Therapeutic exposure variables consisted of 15 measures of pharmacological treatment and management. These included corticosteroids, nonsteroidal anti-inflammatory drugs (regular, on-demand, or none), biologic therapies (infliximab, etanercept, or adalimumab), and conventional synthetic DMARDs (methotrexate, sulfasalazine, or leflunomide).

The disease activity variables included measures of clinical status, functional impairment, and patient-reported outcomes. Disease activity was assessed with the Ankylosing Spondylitis Disease Activity Score (ASDAS). Functional impact and quality of life were measured using the Ankylosing Spondylitis Quality of Life (ASQoL) questionnaire and the physical and mental component scores of the 12-item Short Form (SF-12) survey. Axial and peripheral involvement was evaluated through the number of swollen joints, chest expansion, and Schober test. The dataset also recorded the clinical form of the disease (axial, peripheral, enthesitic, or mixed), extra-articular manifestations (iritis, uveitis, enthesitis, dactylitis, pustulosis, acne conglobata, balanitis, prostatitis, and nail involvement), comorbidities (cardiac, renal, pulmonary, and neurologic), and disease duration (years from symptom onset to the baseline visit). Pain was assessed with the VAS, with which patients rated their overall pain (not restricted to back pain) from 0 (no pain) to 10 (worst imaginable pain) based on the previous week.

Data analysis. We merged REGISPONSER and RESPONDIA and removed outliers and data errors. We summarized numeric and categorical variables to characterize the sample; baseline demographic, clinical, and social characteristics are shown in Table 1. More detailed distributions and missing data summaries by cohort, together with boxplots of VAS across categorical variables and kernel density plots for numerical variables, are provided in Supplementary Files 1 and 2.

Given nontrivial missingness, we used multiple imputation to create m = 20 completed datasets.[14] We rejected Missing Completely at Random via the Little test and proceeded under a Missing-At-Random assumption.[15] To study social determinants of pain, we built an SFI from education, employment/occupation, income source, housing, and the Graffar scale. In each imputation, models used penalization with out-of-fold predictions to avoid optimism; when group lasso was unavailable, we used elastic net with internal cross-validation.[16] Individual SFI scores were averaged across imputations. We evaluated SFI performance via correlation and discrimination for high pain (area under the curve [AUC] for VAS ≥ 7), reporting per imputation and pooled metrics (Figure 1).

To quantify associations with VAS beyond linear effects, we computed MI between VAS and candidate predictors—including SFI—within each imputed dataset, following recent applications in rheumatology.[17–20] We used the standard definition

$$\mathrm{MI}(X, VAS) = H(VAS) - H(VAS|X),$$

where $H(\cdot)$ is the Shannon entropy function. Practically, we discretized VAS into five quantile-based bins, encoded categorical variables, scaled continuous predictors, and estimated MI with bootstrap uncertainty (100 resamples per imputation), then we pooled ranks across imputations (Figure 2). MI guided the choice of pre-exposure covariates for causal analyses.

We estimated the average treatment effect on the treated (ATT) by contrasting high (Q4) and low (Q1) SFI scores using propensity score matching with logistic model scores, overlap trimming,[21] 1:1 nearest-neighbor matching without replacement on the logit scale, and a 0.2 SD caliper (with 0.3 SD as a sensitivity check).[22] Balance was assessed using standardized mean differences (SMDs), and ATT estimates with SEs were pooled across imputations via Rubin rules. We additionally fit doubly robust outcome regressions for covariates with residual imbalance. The pooled ATT (95% confidence intervals [CIs]) is shown in Figure 3 (forest plot), covariate balance before and after matching in Figure 4 (love plot), and paired postmatching VAS differences in Figure 5 (distribution of matched differences). Full specifications, diagnostics, and sensitivity analyses are provided in the Supplementary Methods; the Supplementary Materials also include SFI validation outputs and metrics (Supplementary File 3), full PSM results with per imputation and pooled ATT (Supplementary File 4), detailed balance tables (Supplementary File 5), doubly robust regression outputs (Supplementary File 6), caliper sensitivity (0.2 vs 0.3 SD; Supplementary File 7) , and stratified Spain vs non-Spain analyses (Supplementary File 8).

All patients gave their written consent to participate in the REGISPONSER and RESPONDIA registries. The study was conducted in compliance with the recommendations of the Declaration of Helsinki and was centrally approved by the Ethics Committee of the Reina Sofia University Hospital from Córdoba (Spain) on April 21, 2006 ("Comisión de Ética e Investigación Sanitaria, REGISPON-2004 y RESPONDIA-2006").

RESULTS

In total, 2,042 patients were included (REGISPONSER N = 1,514; RESPONDIA N = 528). A concise summary of baseline demographic, clinical, and social characteristics is provided in Table 1. Patients from REGISPONSER were on average (mean ± SD) older (48.1 ± 12.9 years) than those from RESPONDIA (44.7 ± 14.7 years), and the proportion of men was higher (74.7% vs

Table 1. Baseline characteristics of patients with spondyloarthritis in the joint pooled dataset, REGISPONSER and RESPONDIA cohorts*

Variable	Dataset (N = 2042)	REGISPONSER (n = 1,514)	RESPONDIA (n = 528)
Demographics, mean ± SD			
Age, y	47.3 ± 13.4	48.1 ± 12.9	44.7 ± 14.7
Diagnostic delay, y	7.5 ± 9.3	7.7 ± 9.4	6.8 ± 9.1
Disease duration, y	19.5 ± 13.1	21.1 ± 13.1	14.8 ± 11.6
Symptom duration, y	11.7 ± 10.3	13.2 ± 10.4	7.4 ± 8.5
Age at diagnosis, y	35.4 ± 12.7	34.9 ± 11.9	37.0 ± 14.9
Weight, kg	73.8 ± 14.2	74.2 ± 14.2	72.4 ± 14.2
Height, cm	166.4 ± 9.3	166.7 ± 9.1	165.4 ± 10.0
Clinical measures, mean ± SD			
Chest expansion, cm	3.6 ± 2.1	3.8 ± 2.2	3.3 ± 1.9
Schober test, cm	3.0 ± 1.9	3.0 ± 1.8	3.2 ± 2.1
Finger-to-floor distance, cm	19.0 ± 15.0	18.6 ± 14.2	20.2 ± 16.8
Occiput-to-wall distance, cm	4.5 ± 6.2	4.4 ± 6.0	4.9 ± 6.6
Swollen joint count	0.8 ± 2.9	0.4 ± 1.8	1.9 ± 4.8
VAS (pain 0–10)	3.3 ± 2.3	3.2 ± 2.1	3.7 ± 2.6
ASQoL (0–18; higher = worse)	6.6 ± 5.2	6.4 ± 5.1	7.4 ± 5.3
SF-12 physical component	37.0 ± 7.5	37.2 ± 7.5	36.3 ± 7.5
SF-12 mental component	50.2 ± 5.5	50.6 ± 5.5	49.3 ± 5.5
ASDAS-CRP	2.4 ± 1.1	2.6 ± 1.1	2.0 ± 1.0
Form, n (%)			
Axial	1,233 (60.9)	1,055 (69.8)	178 (34.6)
Mixed	744 (36.7)	433 (28.7)	311 (60.5)
Peripheral	39 (1.9)	19 (1.3)	20 (3.9)
Enthesitic	9 (0.4)	4 (0.3)	5 (1.0)
Family history, n (%)			
No	1,513 (79.5)	1,117 (79.6)	396 (79.2)
Yes	390 (20.5)	286 (20.4)	104 (20.8)
Sex, n (%)			
Male	1,485 (72.7)	1,131 (74.7)	354 (67.0)
Female	557 (27.3)	383 (25.3)	174 (33.0)
Education, n (%)			
High school	606 (41.6)	359 (38.1)	247 (47.9)
Elementary school	556 (38.1)	433 (46.0)	123 (23.8)
Illiterate	20 (1.4)	13 (1.4)	7 (1.4)
University	276 (18.9)	137 (14.5)	139 (26.9)
Marital status, n (%)			
Married	1,023 (70.2)	713 (75.6)	310 (60.2)
Single	364 (25.0)	194 (20.6)	170 (33.0)
Divorced/separated	48 (3.3)	25 (2.7)	23 (4.5)
Widowed	23 (1.6)	11 (1.2)	12 (2.3)
Profession, n (%)			
Unskilled worker	377 (25.9)	238 (25.3)	139 (27.0)
Skilled worker	349 (24.0)	283 (30.1)	66 (12.8)
Employee	309 (21.3)	201 (21.4)	108 (21.0)
Technician	219 (15.1)	129 (13.7)	90 (17.5)
University profession	200 (13.8)	88 (9.4)	112 (21.7)
Source of income, n (%)			
Salary	1,022 (70.3)	643 (68.5)	379 (73.6)
Donations	216 (14.9)	150 (16.0)	66 (12.8)
Profits	113 (7.8)	61 (6.5)	52 (10.1)
Rents	103 (7.1)	85 (9.1)	18 (3.5)
Housing conditions, n (%)			
Good housing	735 (50.5)	463 (49.2)	272 (53.0)
Excellent (no luxury)	588 (40.4)	411 (43.6)	177 (34.5)
Deficient	86 (5.9)	47 (5.0)	39 (7.6)
Excellent with luxury	24 (1.6)	9 (1.0)	15 (2.9)
Inadequate	22 (1.5)	12 (1.3)	10 (1.9)
Graffar scale, n (%)			
Middle-lower class	594 (41.0)	411 (43.9)	183 (35.7)
Middle class	443 (30.6)	288 (30.8)	155 (30.2)
Upper-middle class	285 (19.7)	166 (17.7)	119 (23.2)
Lower class	95 (6.6)	59 (6.3)	36 (7.0)

(Continued)

Table 1. (*Cont'd*)

Variable	Dataset (N = 2042)	REGISPONSER (n = 1,514)	RESPONDIA (n = 528)
Upper class	32 (2.2)	12 (1.3)	20 (3.9)
Race, n (%)			
White	1,193 (81.4)	930 (98.3)	263 (50.6)
White-Indigenous	201 (13.7)	2 (0.2)	199 (38.6)
White-Black	23 (1.6)	3 (0.3)	20 (3.8)
Other	18 (1.2)	7 (0.7)	11 (2.1)
Black	13 (0.9)	1 (0.1)	12 (2.3)
Indigenous	10 (0.7)	1 (0.1)	9 (1.7)
Black-Indigenous	5 (0.3)	1 (0.1)	4 (0.8)
Indigenous-Asian	3 (0.2)	1 (0.1)	2 (0.4)
Employment status, n (%)			
Employed	1,038 (51.6)	738 (49.5)	300 (57.6)
At home	580 (28.8)	525 (35.2)	55 (10.6)
Unemployed	309 (15.4)	182 (12.2)	127 (24.4)
Student	19 (0.9)	5 (0.3)	14 (2.7)
Disability, n (%)			
No disability	1,417 (72.0)	1,047 (71.0)	370 (75.2)
Permanent disability	444 (22.6)	367 (24.9)	77 (15.7)
Temporary disability	106 (5.4)	61 (4.1)	45 (9.1)
Country, n (%)			
Spain	1,514 (74.1)	1,514 (100.0)	0 (0)
Argentina	147 (7.2)	0 (0)	147 (27.8)
Mexico	103 (5.0)	0 (0)	103 (19.5)
Portugal	88 (4.3)	0 (0)	88 (16.7)
Chile	76 (3.7)	0 (0)	76 (14.4)
Venezuela	38 (1.9)	0 (0)	38 (7.2)
Peru	32 (1.6)	0 (0)	32 (6.1)
Uruguay	28 (1.4)	0 (0)	28 (5.3)
Costa Rica	16 (0.8)	0 (0)	16 (3.0)

* Percentages use the variable-specific denominator because of missing data. ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score – C-Reactive Protein; ASQoL, Ankylosing Spondylitis Quality of Life; SF-12, 12-item Short Form; VAS, Visual Analog Scale.

67.0%). Regarding disease activity and pain, the mean ± SD VAS score was 3.20 ± 2.13 in REGISPONSER and 3.72 ± 2.58 in RESPONDIA, whereas the ASDAS was 2.58 ± 1.12 and 2.03 ± 1.01, respectively. Socioeconomic differences were also observed: for example, university education and employment were more common in RESPONDIA. Comparative boxplots of all variables can be consulted in boxplots.xlsx, where it can also be observed that patients in the Graffar upper class reported lower VAS levels. Detailed descriptive statistics (means, SD, quartiles, n and %, and missing data) by registry are provided in descriptive.docx.

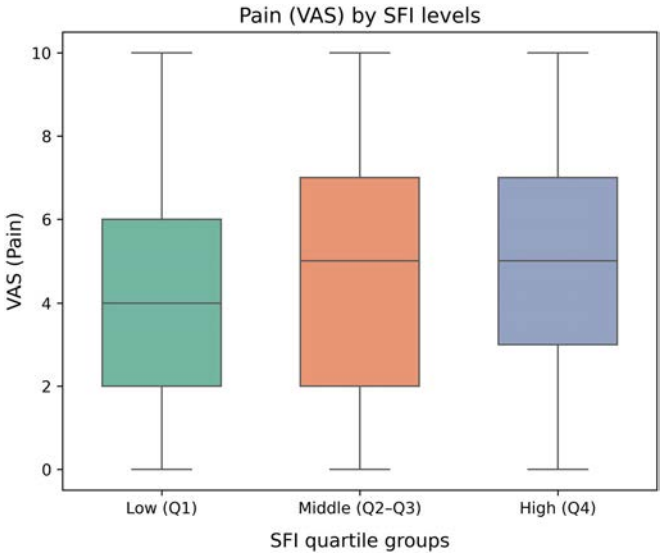


Figure 1. Distribution of pain scores (VAS) across quartiles of the SFI. SFI, Social Factors Index; VAS, Visual Analog Score. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25675/abstract.

The SFI was defined as a weighted sum of nine social factors each compared with the middle baseline of its domain:

$$\begin{aligned} SFI = &+0.0467\,[\textit{Employment Status\_Student}] \\ &+0.0177\,[\textit{Employment Status\_Unemployed}] \\ &-0.0175\,[\textit{Employment Status\_Employed}] \\ &+0.0179\,[\textit{Education\_Illiterate}] - 0.0236\,[\textit{Education\_University}] \\ &+0.0414\,[\textit{Housing Conditions\_Inadequate}] \\ &-0.0224\,[\textit{Housing Conditions\_Deficients}] \\ &+0.0246\,[\textit{Graffar Scale\_Lower Class}] \\ &-0.0211\,[\textit{Graffar Scale\_Upper Class}]. \end{aligned}$$

Positive coefficients indicate higher expected pain (VAS) relative to baseline. The index showed a modest but consistent association with VAS (Pearson $r \approx 0.17$) and modest discrimination for high pain (VAS ≥ 7, AUC ≈ 0.56). However, the boxplot of VAS across SFI quartiles (Figure 1) showed an upward shift in both

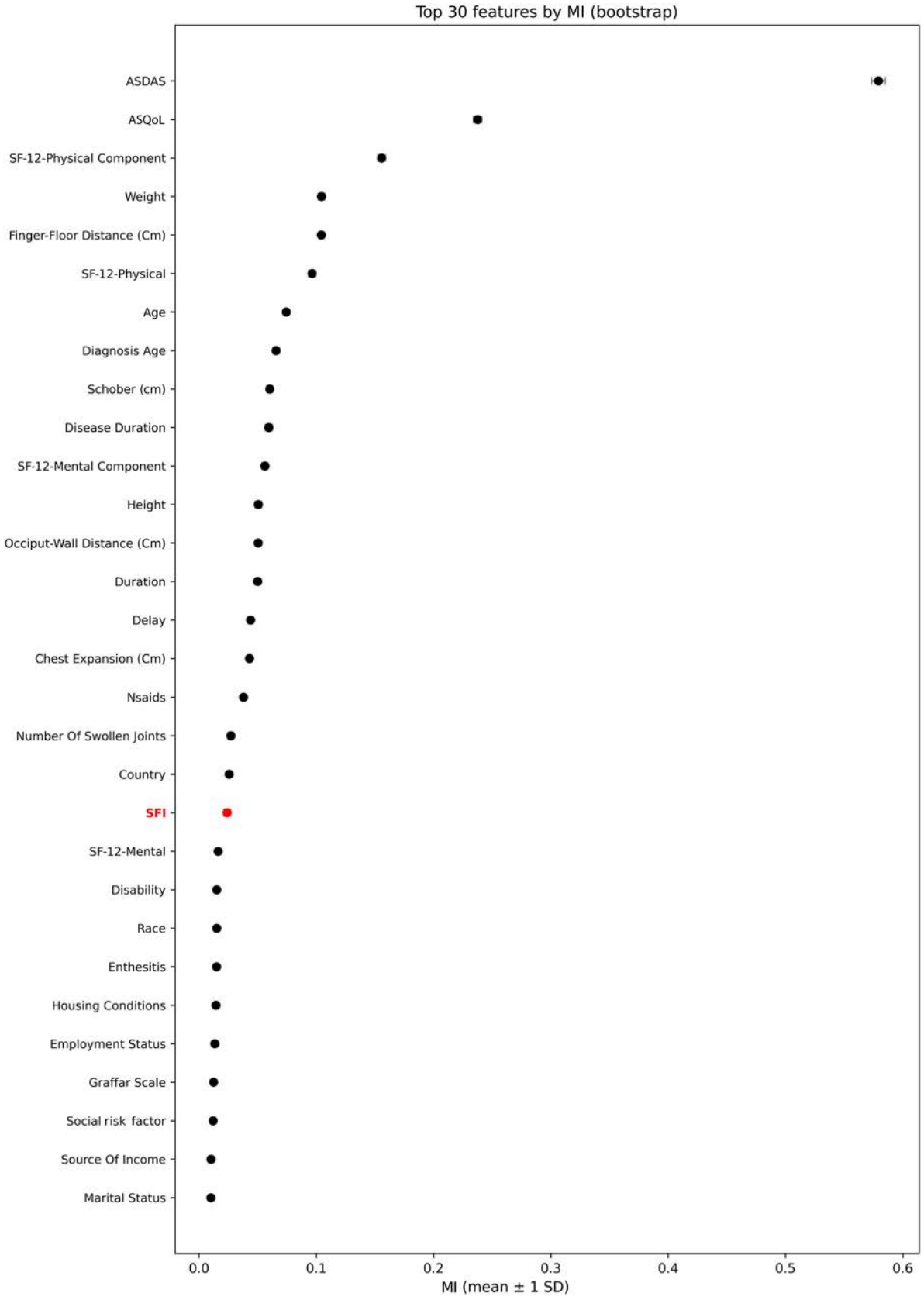


Figure 2. MI with the Visual Analog Scale for the 30 clinical and social features, computed across 100 bootstrap iterations. We highlighted the SFI in red to show its relative position with respect to the other variables. ASDAS, Ankylosing Spondylitis Disease Activity Score; ASQoL, Ankylosing Spondylitis Quality of Life; MI, mutual information; NSAID, nonsteroidal anti-inflammatory drug; SF-12, 12-item Short Form; SFI, Social Factors Index.

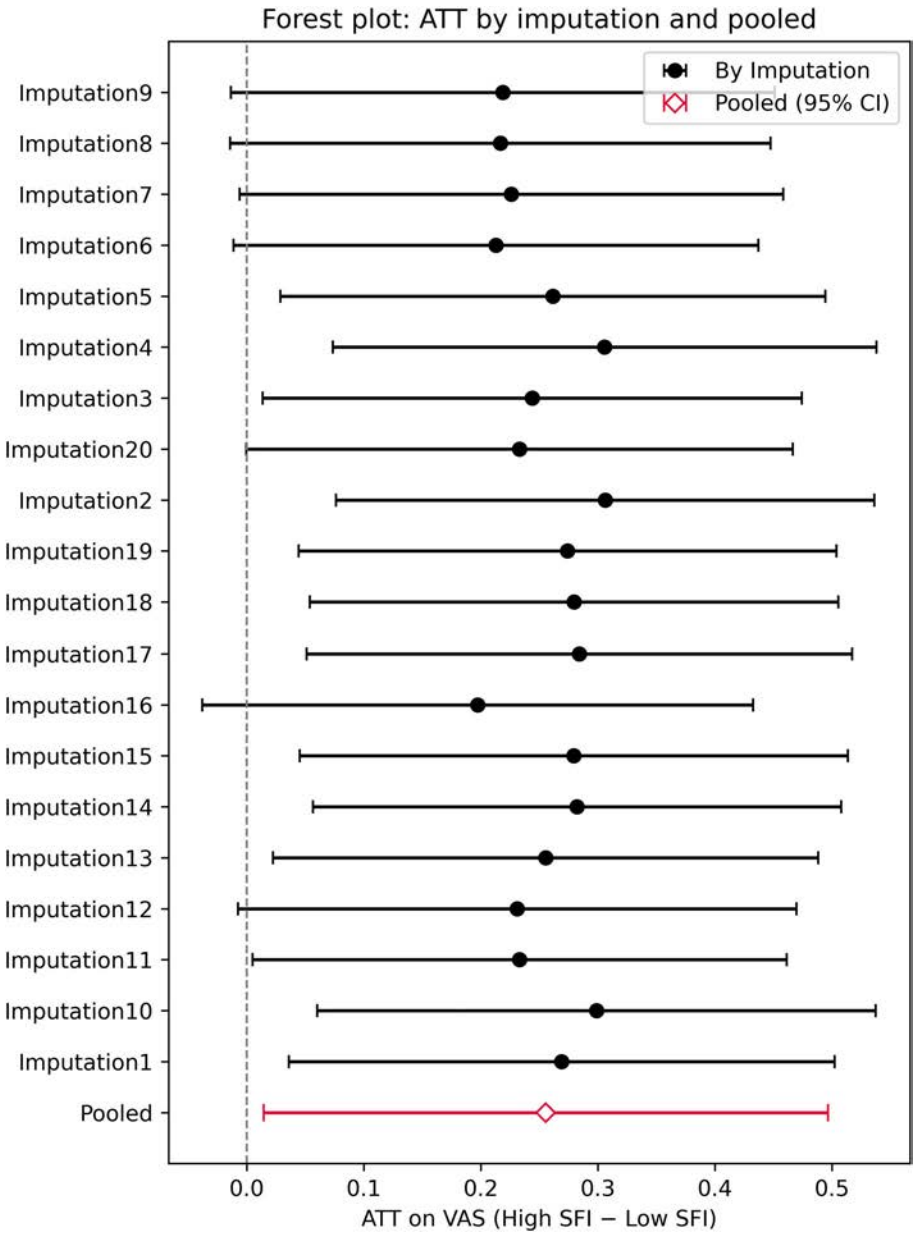


Figure 3. Forest plot of ATT estimates for VAS across imputations and pooled results. ATT, average treatment effect on the treated; CI, confidence interval; SFI, Social Factor Index; VAS, Visual Analog Scale. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25675/abstract.

medians and interquartile ranges, suggesting higher pain levels with increasing SFI.

MI ranking (Figure 2) showed that the SFI captured more information about VAS than any single social variable. The pooled MI for the SFI index was 0.0239, placing it 20th overall and ahead of other social indicators such as the Graffar scale (0.0124), employment status (0.0134), housing conditions (0.0144), source of income (0.0102), race (0.0151), and country (0.0257). As expected, clinical/activity measures dominated the top ranks (eg, ASDAS 0.579, ASQoL 0.237, weight 0.104, and SF-12 physical 0.156). As we described in the previous section, we also used these MI results to select those variables ranked high and plausibly related to both social risk and pain. This led to the final set of baseline covariates: age (0.074), diagnosis age (0.066), disease duration (0.059), diagnostic delay (0.044), weight (0.104), height (0.051), race (0.015), and country (0.026).

Regarding the results of the propensity score matching analysis, we compared pain outcomes between patients with high social factor scores (Q4 of the SFI) and those with low scores (Q1). Before matching, Q1 and Q4 included, on average, 918 and 829 patients per imputation, respectively, reflecting roughly one-quarter of the cohort. The pooled ATT was +0.38 VAS points (SE = 0.17), with a 95% CI of 0.05 to 0.71 and a *P* value of 0.023 across the 20 imputations (Figure 3). After

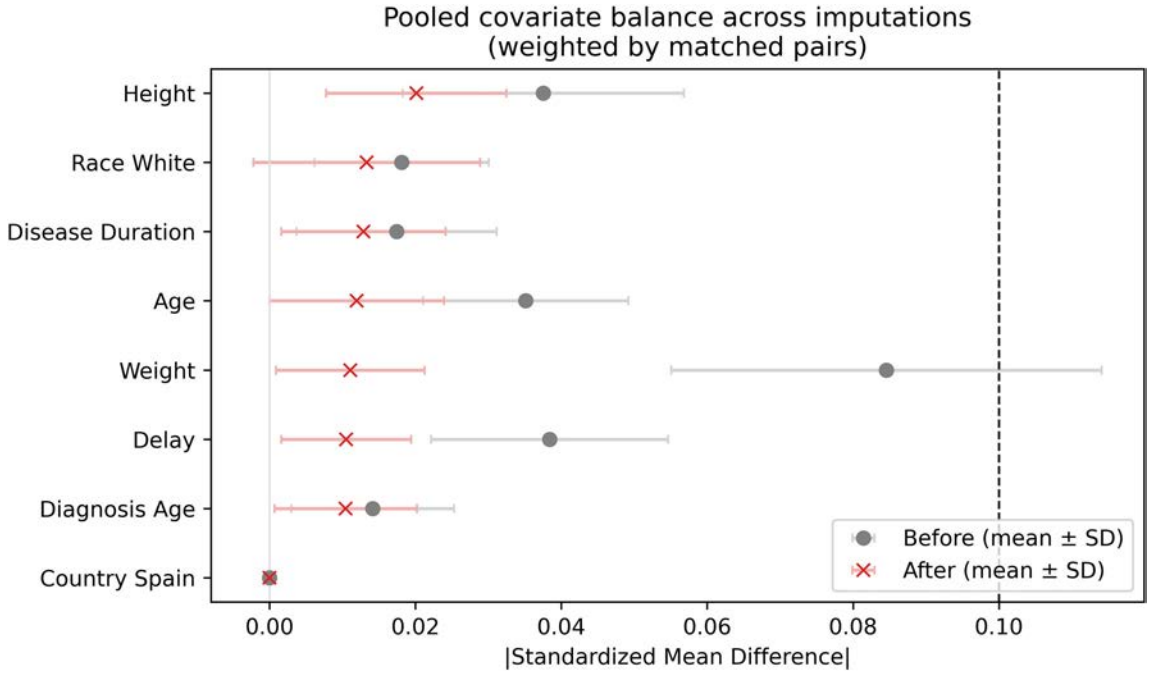


Figure 4. Love plot depicting covariate balance before and after propensity score matching. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25675/abstract.

overlap trimming and 1:1 nearest-neighbor caliper matching, an average of 745 matched pairs per imputation was retained (range 731–761), corresponding to approximately 90% of eligible patients in the extreme quartiles. Matching quality was excellent, with a mean absolute standardized difference of 0.041 across covariates. Figure 4 displays the pooled SMDs across imputations for all variables in the propensity model, weighted by matched pairs. Before matching (gray dots), some covariates showed imbalance, particularly race (SMD = –0.16) and country (SMD = –0.27). After matching, these imbalances were reduced to –0.09 and –0.04, respectively, whereas continuous covariates such as age, diagnosis age, disease duration, delay, weight, and

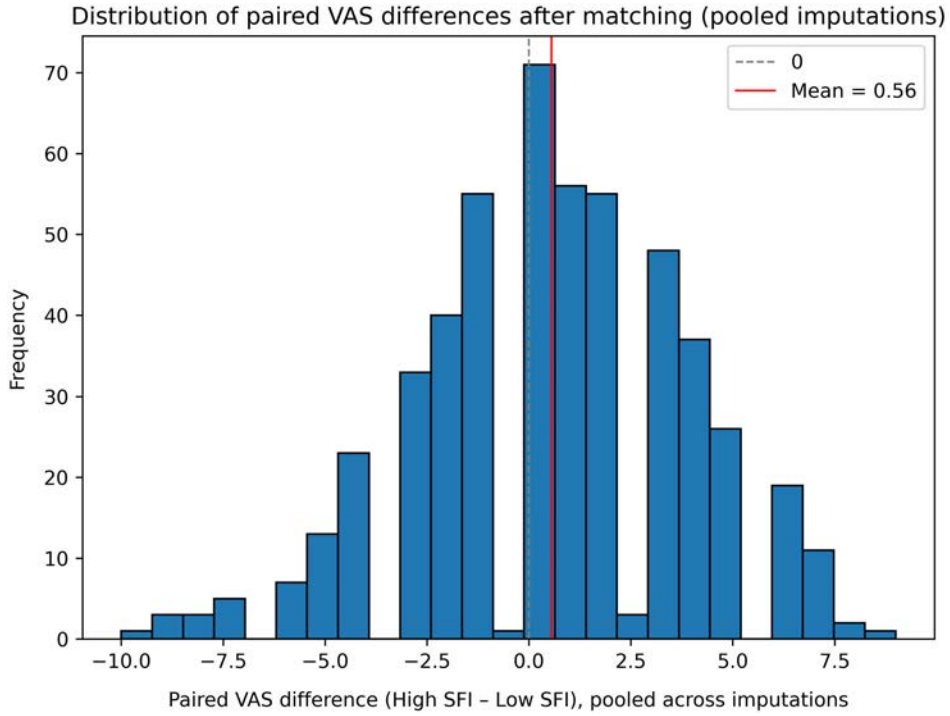


Figure 5. Distribution of paired differences in pain scores after matching. SFI, Social Factors Index; VAS, Visual Analog Scale. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25675/abstract.

height all achieved postmatch SMDs between –0.07 and 0.04. Thus, every covariate met the conventional threshold of |SMD| <0.10. In the matched sample, the distribution of pairwise differences in VAS remained shifted to the right, with a pooled mean difference of +0.38 points (95% CI 0.05–0.71), confirming that pain severity was consistently higher in the high SFI group.

The results of the caliper sensitivity analysis (see caliper_sensitivity.xlsx) were highly consistent across specifications. When widening the caliper to 0.3 SD, results were almost identical to those obtained with 0.2 SD (ATT = +0.41; SE = 0.16; 95% CI 0.09–0.72; P = 0.011), with a similar number of matched pairs (802) and a mean absolute SMD after matching of 0.022. Finally, in the stratified analysis to observe differences between Spain and Latin American countries, we obtained an ATT of +0.31 (SE = 0.20; 95% CI 0.08–0.71) across 20 imputations for Spain and +0.62 (SE = 0.43; 95% CI –0.22 to 1.45) across 20 imputations for Latin America.

DISCUSSION

Some previous studies have primarily attributed the impact of socioeconomic factors not to pain, but to disease activity. For example, Capelusnik et al[7] suggest that lower education levels and single marital status were associated with higher ASDAS. Other studies[23,24] found that lower levels of education and income were associated with a higher likelihood of receiving a chronic pain diagnosis.

Our SFI shows a clear gradient: students and unemployed individuals, those living in inadequate housing, and patients from lower social classes have higher predicted pain, whereas those with university education and from upper social classes have lower predicted pain. These patterns align with evidence that social determinants are linked to worse outcomes in axial SpA.[7] Mendelian randomization work indicates that higher education attainment causally reduces multisite chronic pain risk, plausibly via healthier behaviors and greater psychologic resources.[25] Daily diary research also shows that financial hardship and unemployment amplify day-to-day pain reactivity, particularly among women with chronic musculoskeletal disease who worry about finances when unemployed.[26] Taken together, this supports aggregating correlated social exposures to reflect cumulative disadvantages.

Despite the modest performance of the SFI in predicting VAS (Pearson r ≈ 0.17; AUC ≈ 0.56), a low overall correlation does not rule out meaningful group differences. When we group the SFI into quartiles, Figure 1 shows that median VAS (and its spread) increases stepwise from Q1 to Q4. In addition, the SFI ranked first among all social indicators in the MI analysis, ahead of education, Graffar scale, employment status, housing conditions, and source of income (Figure 2). This means that, although social context explains less variance in pain than clinical activity, the combined index captures more information about VAS than any single social variable, supporting its use as a compact summary of cumulative social factors.

Using propensity score matching to compare patients in the highest SFI quartile (Q4) with those in the lowest (Q1), the pooled ATT for VAS was +0.38 points (SE = 0.17; 95% CI 0.05–0.71; P = 0.023) across 20 imputations (Figure 3). Covariate balance was excellent after matching (mean absolute SMD ≈ 0.041) with approximately 745 matched pairs per imputation (Figures 4 and 5), and results were stable to caliper changes (0.2–0.3 SD), consistent with prior methodologic work.[27–29] In plain terms, patients with a high SFI score reported approximately 0.4 points more VAS pain than comparable patients with a low SFI score. Although modest, this effect was consistent and clinically relevant. The gradient was more pronounced in Latin America (ATT = +0.62; 95% CI –0.22 to 1.45) than in Spain (ATT = +0.31; 95% CI –0.08 to 0.71), suggesting stronger social disparities in the former.

Our methodology has several strengths. We used multiple imputations (20 datasets) and derived a stable, interpretable SFI via sparse-group lasso with bootstrap resampling. MI was applied only to assess the SFI's informativeness for VAS, not for confounder selection. Prespecified baseline covariates (age, age at diagnosis, disease duration, diagnostic delay, weight, height, race, and country) achieved excellent postmatch balance (mean absolute SMD ≈ 0.041). Propensity scores were estimated with logistic regression, prioritizing balance over prediction; effects were robust to caliper variation on the logit scale (0.2–0.3 SD) and were larger in Latin America than in Spain.

However, several limitations should be noted. First, the study is observational and cross-sectional rather than longitudinal, which limits the ability to establish causal relationships and raises the possibility of reverse causation. Second, although we adjusted for a wide range of demographic and clinical covariates, residual confounding cannot be excluded, particularly from unmeasured psychosocial factors such as depression, coping styles, health literacy, or differences in health care access. Third, the SFI was developed primarily in a Spanish cohort, and although the effect appeared stronger in Latin America, where social inequalities are more pronounced, external validation and larger datasets from these countries are needed to confirm and generalize these findings.

This study suggests that social factors could contribute independently to pain severity in patients with SpA. We summarized education, employment status, Graffar scale, and housing conditions into a single SFI, showing a clear gradient in pain: VAS increased stepwise from the lowest to the highest SFI quartile. With propensity score matching, patients in the highest SFI quartile had modest yet consistent increases in VAS compared with otherwise comparable patients in the lowest quartile (ATT = +0.38; 95% CI 0.05–0.71; P = 0.023), supporting an effect beyond measured disease activity. Although the individual-level effect size is small, such gradients can be meaningful at the population level and have policy relevance, helping identify groups who may benefit from targeted support. Clinically, patients with

lower education, unstable employment, or inadequate housing may benefit from targeted assessment and support (eg, pain self-management support or social work referral). Future work should externally validate the SFI and matched findings, use longitudinal designs to establish temporality, and test interventions that address the identified social pathways.

ACKNOWLEDGMENTS

The authors would like to thank all the investigators from the REGISPONSER and RESPONDIA study groups. We would also like to acknowledge the support of technological tools such as ChatGPT, whose assistance helped improve the fluency and clarity of our English language expression in some sections of this document.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Mease PJ, Bhutani MK, Hass S, et al. Comparison of clinical manifestations in rheumatoid arthritis vs. spondyloarthritis: a systematic literature review. Rheumatol Ther 2022;9(2):331–378.
2. Lee YC, Malfait AM, Ogdie AR. Unmet needs in spondyloarthritis: understanding and managing chronic pain. J Rheumatol 2024; 51(12):1235–1240.
3. Kiltz U, Baraliakos X, Regel A, et al. Causes of pain in patients with axial spondyloarthritis. Clin Exp Rheumatol 2017;35(5)(suppl 107): 102–107.
4. Pathan EMI, Inman RD. Pain in spondyloarthritis: a neuro-immune interaction. Best Pract Res Clin Rheumatol 2017;31(6):830–845.
5. Capelusnik D, Boonen A, Ramiro S, et al. The role of social determinants of health on disease outcomes in axial spondyloarthritis: a narrative review. Autoimmun Rev 2025;24(5):103762.
6. Santos H, Henriques AR, Branco J, et al. Health-related quality of life among spondyloarthritis and chronic low back pain patients: results from a nationwide population-based survey. Qual Life Res 2023; 32(2):383–399.
7. Capelusnik D, Zhao SS, Boonen A, et al. Individual-level and country-level socio-economic factors and health outcomes in spondyloarthritis: analysis of the ASAS-perSpA study. Rheumatology (Oxford) 2022;61(5):2043–2053.
8. Ruiz-Vilchez D, Ladehesa-Pineda L, Puche-Larrubia MÁ, et al. The socioeconomic status of patients with ankylosing spondylitis and its association with the burden of the disease and permanent disability: a cross-sectional cluster analysis. Ther Adv Musculoskelet Dis 2024; 16:1759720X241272947.
9. Cheung JPY, Cheung PWH, Wong CKH, et al. Propensity-matched comparison between chronic nonspecific low back pain and axial spondyloarthritis: impact on patient-perceived quality of life. Spine 2023;48(8):577–589.
10. Dougados M, van der Linden S, Juhlin R, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34(10):1218–1227. https://doi.org/10.1002/art.1780341003
11. Collantes E, Zarco P, Muñoz E, et al. Disease pattern of spondyloarthropathies in Spain: description of the first national registry (REGISPONSER) extended report. Rheumatology (Oxford) 2007; 46(8):1309–1315.
12. Vazquez-Mellado J, Font Ugalde P, Muñoz Gomariz E, et al. Ibero-American Spondylarthritis Registry (RESPONDIA): what is, how it came about, who we are and what we do. Reumatol Clin 2008;4: S17–S22.
13. Graffar M. Une méthode de classification sociale d'échantillons de population. Article in French. Courrier 1956;6:455–459.
14. Rubin DB. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons; 1987.
15. Little RJA. A test of Missing Completely at Random for multivariate data with missing values. J Am Stat Assoc 1988;83(404):1198–1202.
16. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1): 1–22.
17. Castro Corredor D, Calvo Pascual LÁ, Collantes-Estévez E, et al. Inflammatory activity levels on patients with anti-TNF therapy: most important factors and a decision tree model based on REGISPONSER and RESPONDIA registries. Ther Adv Musculoskelet Dis 2025; 17:X251332224.
18. Calvo Pascual LÁ, Castro Corredor D, Garrido Merchán EC. Machine learning classification of vitamin D levels in spondyloarthritis patients. Intell Based Med 2024;9:100125.
19. Castro Corredor D, Calvo Pascual LÁ. Imbalanced machine learning classification models for removal biosimilar drugs and increased activity in patients with rheumatic diseases. PLoS One 2023;18(11): e0291891.
20. Curtis JR, Su Y, Black S, et al. Machine learning applied to patient reported outcomes to classify physician derived measures of rheumatoid arthritis disease activity. ACR Open Rheumatol 2022;4(12):995–1003.
21. Crump RK, Hotz VJ, Imbens GW, et al. Dealing with limited overlap in estimation of average treatment effects. Biometrika 2009;96(1): 187–199.
22. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009;28(25):3083–3107.
23. Jöud A, Petersson IF, Jordan KP, et al. Socioeconomic status and the risk for being diagnosed with spondyloarthritis and chronic pain: a nested case-control study. Rheumatol Int 2014;34(9):1291–1298.
24. Prego-Domínguez J, Khazaeipour Z, Mallah N, et al. Socioeconomic status and occurrence of chronic pain: a meta-analysis. Rheumatology (Oxford) 2021;60(3):1091–1105.
25. Wang H, Liu M, Li H, et al. Association between educational attainment and chronic pain: a mediation Mendelian randomization study. J Pain Res 2025;18:1793–1804.
26. Rios R, Zautra AJ. Socioeconomic disparities in pain: the role of economic hardship and daily financial worry. Health Psychol 2011;30(1): 58–66.
27. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Stat Med 2010;29(3):337–346.
28. Pirracchio R, Petersen ML, van der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. Am J Epidemiol 2015;181(2):108–119.
29. Zhao P, Su X, Ge T, et al. Propensity score and proximity matching using random forest. Contemp Clin Trials 2016;47:85–92.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 106–112
DOI 10.1002/acr.25660

American College of Rheumatology
Empowering Rheumatology Professionals

Rethinking Strategies for a Pharmaceutical Approach to Pain Related to Connective Tissue–Related Raynaud Phenomenon in the United States

Tracy M. Frech,[1] Charles G. Frech,[2] W. David Merryman,[2] Andrew Sternlicht,[3] and Justin Baba[2]

Objective. There are no US Food and Drug Administration–approved therapies for Raynaud phenomenon (RP) in the United States. Clinical trials have been challenged by study design. Important advances in RP patient-reported outcome measures and mechanistic quantification allow RP-related pain characterization. The rationale for this narrative review is current RP treatment guidelines that focus on vasodilation.

Methods. The question of why there are limitations to RP treatment in the United States is addressed through a comprehensive search strategy of published RP treatment guidelines up until September 1, 2025. Search databases included Medline (PubMed), Embase, and Scopus for the index terms "Raynaud's phenomenon treatment guidelines." If a society guideline was updated, only the most recent was included. Eligibility, data extraction, risk of bias, and quality assessment were subject to review by two independent reviewers, with a third reviewer resolving discrepancies. US-specific considerations of published guidelines are reviewed.

Results. A total of 118 published articles were identified by the search terms "Raynaud's phenomenon treatment guidelines," and 27 abstracts were reviewed. Four articles published as RP treatment recommendations or guidelines were reviewed for full content. Pain management for RP is not included in guideline-based care.

Conclusion. There are advances in outcome measures for quantifying pain now available for RP clinical trials. Large US-based registries for systemic sclerosis using patient-reported outcomes can allow serial data collection on RP and RP-related digital lesions to provide real-world data on medication efficacy for pain relief.

Introduction

Raynaud phenomenon (RP) is a common painful neurosensory symptom related to digital (finger and toes) vasospasm in response to cold temperature and emotional stress, and it has a prevalence of approximately 8% in the United States.[1] RP is a symptom complex related to digital vascular compromise, which always includes blanching triggered by cold exposure but is associated with various conditions with different outcomes.[2] Primary RP, although symptomatically bothersome, is not associated with vascular damage or autoantibodies,[3] whereas secondary RP, related to a connective tissue disease (CTD), typically exhibits abnormal nailfold microscopy and in its severe form, most commonly related to systemic sclerosis (SSc), results in vasculopathy-driven digital lesions. RP is often the earliest symptom of autoimmune disease and is diagnosed by performing a physical examination of the hand, specifically by assessing digital pitting, ulceration, and telangiectasias and performing nailfold capillaroscopy. History and physical examinations can be helpful in distinguishing painful mimics of RP. Ancillary studies such as hand radiographs can identify severe autoimmune disease features such as digital tip resorption (acro-osteolysis) and calcinosis, which are thought to be a tissue response to sustained or prolonged hypoxia and possibly prevented with vasodilator therapy.[4] If results of the physical examination or ancillary studies are consistent with RP and abnormal, laboratory testing can identify CTD autoantibodies. In fact, a patient with RP, digital lesions (pits and/or ulcers), abnormal capillaroscopy, and an SSc autoantibody meets classification criteria for this CTD, which has among the highest morbidity and mortality of all rheumatic diseases.[5] Thus, it is critically important to standardize the screening and treatment of RP, specifically if treatments are preventive.[6] Ethnic, cultural, and clinical phenotypic factors appear to contribute to geographic variation in RP symptom burden in SSc,

Dr T. M. Frech's work was supported by a Department of Veterans Affairs Merit Award (grant I01CX002111).

[1]Vanderbilt University Medical Center and Tennessee Valley Healthcare System, Nashville, Tennessee; [2]Vanderbilt University, Nashville, Tennessee; [3]Tufts University School of Medicine, Boston, Massachusetts.

Author disclosures and graphical abstract are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25660.

Address correspondence via email to Tracy M. Frech, MD, MSCI, at tracy.frech@vumc.org.

Submitted for publication July 8, 2025; accepted in revised form September 16, 2025.

SIGNIFICANCE & INNOVATIONS

- Recent significant advances in the development of patient-reported outcomes, which are grounded in the patient experience, allow for proper quantification of Raynaud phenomenon (RP) in systemic sclerosis and can allow for understanding of longitudinal patterns of pain and pain transitions related to RP.
- The use of vasodilators for pain management in the United States is significantly impacted by access to off-label therapies not approved by the US Food and Drug Administration for this indication.
- There are innovative new assessment tools, such as automated interpretation of capillaroscopy reports and remote training, that may assist in improved diagnosis.
- In this review article, new mechanistic insights for pain management with vasodilator treatment of RP are discussed.

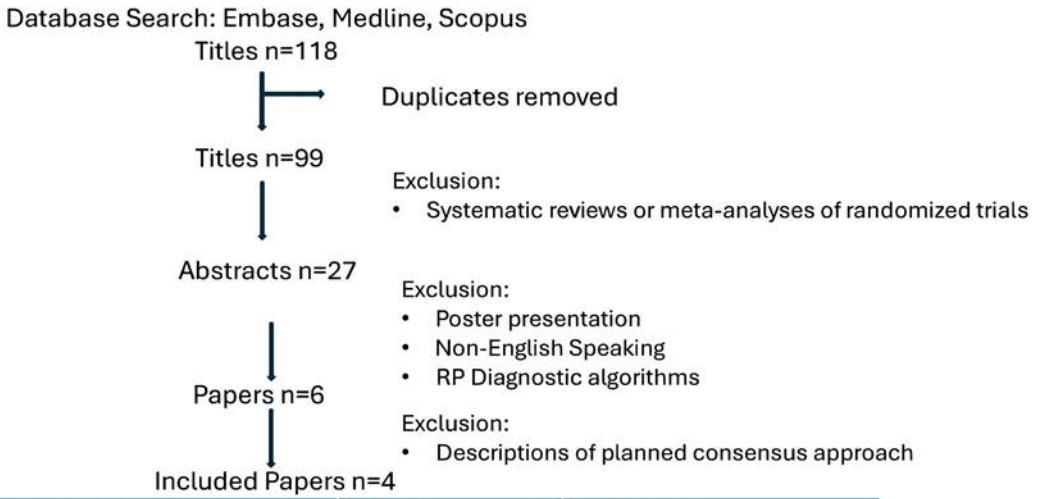
highlighting the value of regional studies.[7] This narrative review addresses the challenges of treatment and opportunities for understanding longitudinal patterns of pain and pain transitions related to RP in the United States. The rationale for this review is that the pain component of RP is not included in current treatment guidelines that focus on vasodilation, and advances in outcome measures can inform a state-of-the-art approach to studying this symptom.

Methods

A comprehensive search of published RP treatment guidelines up until September 1, 2025, was performed. Search databases included Medline (PubMed), Embase, and Scopus for the index terms "Raynaud's phenomenon treatment guidelines" OR (Raynauds AND phenomenon AND treat/exp OR treatment) AND (guidelines/exp OR guidelines). Articles that were diagnostic guidelines, treatment systematic reviews or meta-analyses of randomized trials, poster presentations, and non-English guidelines were excluded. If a society guideline was updated, only the most recent was included. Eligibility, data extraction, and quality assessment were subject to review by two independent reviewers (TMF and CGF), with a third reviewer resolving discrepancies (JB).

Results

As shown in Figure 1, 118 published articles were identified by the search terms "Raynaud's phenomenon treatment guidelines," and 27 abstracts were reviewed. One potentially eligible abstract from the Chinese Rheumatology Association was excluded as



EULAR	SSc Expert Consensus	ESVM	BSR/BHPR
CCB	CCB	CCB	CCB
Prostanoid	PDE5I	ARB	ARB
PDE5I	ARB	SSRI	SSRI, alpha blockers, ACE inhibitors, and statins
SSRI	Prostanoid	Prostanoid	Prostanoid

Figure 1. Narrative review of published treatment guidelines for RP. ACE, acetylcholine esterase; ARB, angiotensin receptor blocker; BSR/BHPR, British Society for Rheumatology/British Health Professionals in Rheumatology; CCB, calcium channel blocker; ESVM, European Society for Vascular Medicine; PDE5I, phosphodiesterase 5 inhibitor; RP, Raynaud phenomenon; SSc, systemic sclerosis; SSRI, selective serotonin reuptake inhibitor. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25660/abstract.

non-English[8]; others were diagnostic (not treatment) recommendations or a poster. Four articles were published as RP treatment recommendations or guidelines. These articles included the EULAR 2023 recommendations for the treatment of SSc,[9] the European Society for Vascular Medicine 2017 recommendations for the diagnosis and management of RP,[2] 2018 treatment algorithms by consensus of SSc experts,[10] and the British Society for Rheumatology/British Health Professionals in Rheumatology 2016 guideline for treatment of SSc[11] (Figure 2). Pain management for RP is not included in any of the guideline-based recommendations. Intravenous prostanoid therapy for RP is in all published guidelines but is cost-prohibitive in the United States for routine use.

What is already known about this subject?

There are guidelines for the treatment of RP[2,9–11] designed to escalate therapies for severity. The guidelines focus on CTD-related RP and include use of prostanoids, which is not covered by insurance companies in the United States and is thus cost-prohibitive. Although patient education and lifestyle adaptations are first-line, guideline-based treatments for RP, access to the wide range of pharmacological options outlined in international guideline-based care for RP can be difficult in the United States because there are no US Food and Drug Administration (FDA)–approved drugs for this indication. Although treatment of RP aims to minimize the occurrence of painful episodes and prevent hand-related disability, the medication management in guidelines focuses on vasodilators. It is also noted that some medications are associated with exacerbation of RP and thus should be stopped if digital symptoms worsen.[12,13] These medications can include stimulants taken for attention deficit disorder, certain migraine medications, and beta-blockers. A medication review is a critical aspect of clinical trial exclusion criteria for RP.

The FDA requires pharmaceutical intervention to achieve significance in impacting feel, function, or survival for approval. The regulatory authorities in the United Kingdom (Medicines and Healthcare products Regulatory Agency), the European Union (European Medicines Agency), and Japan (Pharmaceuticals and Medical Devices Agency) have approved the use of calcium channel blockers (CCBs), phosphodiesterase 5 inhibitors (PDE5-I), endothelial receptor antagonists, and prostacyclin analogues for the treatment of RP related to SSc. Since1980 there have been several trials designed to test treatment of RP and its severe manifestation of digital ulceration (DU). Key aspects of previous studies include identification that a clinical trial length should be specific to cold-weather months to capture drug effect, global representation of study participants but regional subanalysis to capture environmental effects, and stratification of RP severity, including failed prior therapies and disease duration of SSc. A critical component of outcome measures that are used in clinical trials is the ability to quantify disease activity versus prevention or occurrence of irreversible damage.

Nailfold capillaroscopy is used diagnostically to determine if RP is related to CTD. The importance of this diagnostic test highlights that RP secondary to CTD can result in tissue damage, such as DU. An abnormal capillaroscopy result in the setting of SSc-specific antibodies has a high probability of the patient meeting classification criteria in the upcoming five years. Thus, this screening procedure is critical for rheumatologists to perform. However, there are no formal US-based training programs. Historically, this is because capillaroscopy is not a billable procedure

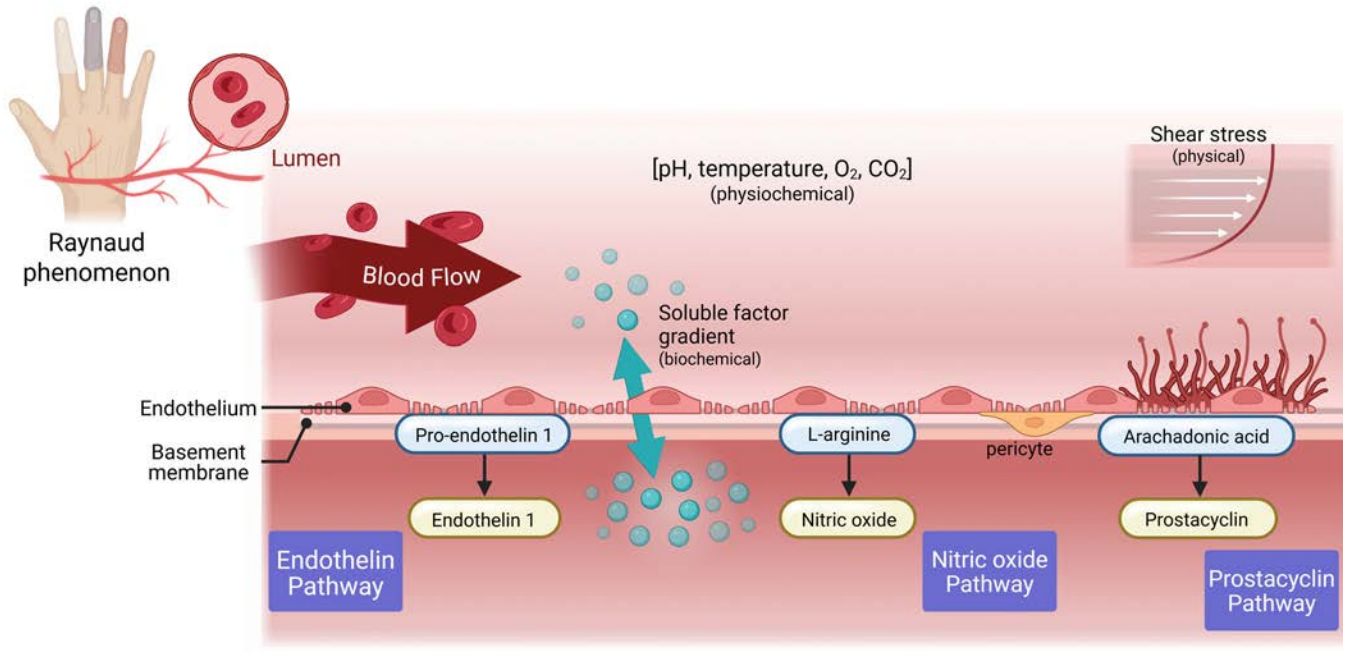


Figure 2. Mechanism of action should address the role of barrier function of vascular endothelium with connective tissue.

to insurance companies, and the gold standard equipment to perform capillaroscopy is expensive and time consuming to use in clinical practice. Advances in low-cost image acquisition with automated interpretation, as well as online training programs, have made this critical diagnostic tool feasible. Nailfold capillaroscopy also may be helpful for RP clinical trial enrichment, as capillary density dropout is strongly associated with DU, which is a painful end-stage manifestation. The correlation of other nailfold abnormalities, such as the presence of microhemorrhage and abnormal size and shape of the vessels, with pain is less clear. Nonetheless, with the reduction of costs of devices that capture nailfold images, online training offered by the EULAR Study Group on Microcirculation in Rheumatic Diseases, and decentralized capillaroscopy interpretation offered by cloud-based systems,[14,15] US-based rheumatologists will be able to use nailfold capillaroscopy reliably. The lack of nailfold videocapillaroscopy as a diagnostic and prognostic tool in clinical practice has potentially impeded US-based RP studies.

What does this study add?

There are advances in outcome measures that capture pain, which can be used in RP clinical trials. The concept of how a therapeutic intervention impacts how a patient feels is a critical aspect of FDA regulatory approval and underscores the importance of understanding RP-related pain. Attack rates, severity scores, participant-preference scores, and physiologic measurements have underperformed in randomized control trials (RCTs).[16,17] Furthermore, the presence of pain related to RP may impact a patient's willingness to take therapeutics or enroll in a clinical trial with this symptom or a resultant DU as the primary outcome, adding selection bias to possible RCT design. As such, easy-to-use instruments to measure RP must be sensitive to change, reliable, valid, and grounded in the patient experience.[18] The lack of patient involvement in the conceptual framework, domain generation, item generation, cognitive interviewing, and respondent burden has been the main criticism of the Raynaud's Condition Score (RCS), which is the most widely used instrument in clinical trials studying moderate to severe RP.[19] The RCS uses a 0 to 10 Likert scale for patients to document in a daily diary their assessment of the combination of severity, frequency, duration, and impact of their RP for that day.[20] The reporting value is the mean of two weeks of daily ratings; thus, as a composite assessment, the RCS does not adequately capture the potential shorter-term impact of drug effect on symptoms, which has critical importance of assessment of "feel" for FDA drug approval.[21,22] When the RCS diary threshold was applied to patients with SSc in a real-world population, it failed to meet all three criteria of feel, function, and quality of life, highlighting the inability to use this outcome measure to enrich RP clinical trials for regulatory approval.[23,24] An additional concern is that RP outcome measures that incorporate a patient-perceived functional status can be impacted by coping strategies independent of prescribed therapeutics, which highlights that a focus on measures that capture impact on pain is a critical aspect of RP severity assessments.[25,26] Furthermore, RP clinical management and RP intervention trial designs should consider temperature patterns.[27]

RP outcome measures that capture how patients feel and function might prove superior to diary-based approaches for assessing condition severity. An international multicenter validation study of the 27-item Assessment of Systemic Sclerosis-Associated Raynaud's Phenomenon (ASRAP) and the 10-item short form (ASRAP-SF) showed the questionnaires are valid and reliable novel patient-reported outcome measures for assessing the severity and impact of SSc-RP.[28,29] The ASRAP questionnaires had good correlation with instruments for assessing disability, hand function, global health assessment, and perhaps most importantly, if used for general RP assessments, intensity of pain. The benefit of this instrument is that it was developed with patient partners with SSc. Its use in other CTD-related RP or primary RP is not proven. As such, pain numerical rating scales or diaries may provide better measurement if used in all RP subsets regardless of etiology.

Although patient reports can capture RP symptom characteristics, objective assessment of digital microvascular function and morphology is helpful for understanding potential mechanisms of disease. Measurement of the temperature of the digits and thermal gradient (temperature of digits minus temperature of dorsum of the hand) may help discern treatment effect in RP.[30] In conjunction with nailfold capillaroscopy, point-of-care thermographic imaging is a noninvasive quantitative tool for assessing peripheral vasculopathy.[31] Functional microcirculatory changes are evident in both primary RP and RP-related to CTD, thus an important aspect of assessment; however, it is noted that pain and paresthesia are more common in RP related to CTD.[32,33] This distinction of painful blanching of the skin guiding prescribed therapies is important contextually for understanding guideline-based care for RP treatment.[2,9,34]

How might this impact clinical practice or future developments?

Including pain-related mechanisms for RP allows expansion of studies on the mechanism of action of therapies prescribed for RP. RP is considered a disorder of vascular thermoregulatory control. In non–CTD-related RP, in which nailfold capillaroscopy is normal, vasospasm of the digital and cutaneous vessels is believed to occur because of an increased α_{2C}-adrenergic response located on vascular prejunctional terminals and smooth muscle and does not result in vascular pathology.[35] As such, a drug targeting vascular thermoregulatory control mechanisms is important for RP symptomatic modification. In CTD-related RP, in which vascular structural and functional changes, combination therapies which address symptoms and vascular damage may be needed.

Preclinical models of CTD-related RP and SSc delineate the relative contributions of specific ligands, receptors, their signaling

pathways, and feedback mechanisms and have been challenged by lack of fidelity for features of vasculopathy, fibrosis, and autoimmunity. Skin biopsies support the importance of abnormal endothelium and microvascular pericytes in CTD-related RP, which are not present in primary RP. Unbiased spatial proteomics with single-cell resolution in skin biopsy tissues further support the role of the endothelial cell in patients with SSc that expresses markers for endothelial-to-mesenchymal transition and is located in close proximity to immune cells and myofibroblasts.[36] Specifically, pericytes are perivascular mesenchymal stem cells with macrophage-like properties embedded within the endothelial basement membrane and are found primarily around blood capillaries, precapillary arterioles, postcapillary venules, and collecting venules to facilitate and assimilate cell communication. The neurotransmitter noradrenaline, released by sympathetic nerve endings, activated α_2-adrenoceptor on pericytes, which leads to vessel constriction.[37] The potential of vascular therapeutics that can target the pericyte is intriguing. It is critically important to understand the role of immune-mediated vascular dysregulation in CTD-related RP pain.

Therapeutics for RP can target the vascular lumen, barrier function, smooth muscle, or surrounding connective tissue (Figure 2). The wall-to-lumen ratio of the radial artery in primary RP and SSc-related RP is reduced and functionally impaired.[38,39] Treatment with an essential cofactor for endothelial nitric oxide (NO) can improve endothelial function in SSc[40,41]; however, accumulating evidence supports that the interplay between carbon monoxide, generated by heme oxygenase and NO plays a crucial role in vascular homeostasis and regeneration by improving endothelial cell communication with neighboring cells, including smooth muscle cells, immune cells, and pericytes.[42] Nifedipine, a first-line therapy and one of the most widely used pure L-type CCBs for treatment of patients with RP, is hypothesized to reduce osteoblastic differentiation of pericytes, thus reducing vascular calcification that may be important for calcinosis. Of note, cilnidipine is a newer fourth-generation CCB with both L-type and N-type Ca^{2+} channel blockade that has renoprotective, cardioprotective, and neuroprotective effects through direct effect on sympathetic nerve endings.[43] Cilnidipine has potentially clinically relevant activity at Nav1.7, a target in pain treatment, and is effective as a combination therapy with PDE5-I.[44] Cilnidipine is currently under investigation for SSc-RP efficacy, safety, and SSc-RP–related pain.

In severe RP or failing first-line CCB treatment, prescription of a PDE5-I, which protects cGMP from degradation-mediating vascular smooth muscle relaxation through NO signaling, is considered second-line therapy for RP.[45] Of interest, the mechanism of action of pentoxifylline, which can be consider in RP, may be related to phosphodiesterase inhibition in addition to platelet sensitivity to the antiaggregatory action of endogenous prostaglandin I_2, which is an unstable cyclooxygenase metabolite detected first in vascular endothelial cells and is associated with the amelioration of pericyte reduction and the transition to myofibroblasts.[46] Although PDE4-I has not been specifically studied in RP, the PDE4 subfamily (PDE4A, PDE4B, PDE4C, and PDE4D) selectively degrades cAMP and plays a vital role in regulating the balance of second messengers in various tissues.[47] PDE4 is the major subtype of PDE enzymes expressed in immune and inflammatory cells and is under investigation in SSc interstitial lung disease for the potential effect of reversibility of vascular-mediated fibrosis.[48]

Prostacyclin analogues, most commonly intravenous iloprost, are reserved for severe RP, usually after CCB and PDE5-I treatments fail, but have cost-prohibitive prescription restrictions in the United States when used for the primary indication of RP.[45] Fluoxetine, a serotonin reuptake inhibitor, is a cost-effective alternative treatment for mild RP when blood pressure side effects limit vasodilators. Experimental data suggest that serotonin drives fibrosis in the skin and visceral organs, promotes platelet aggregation, induces vasoconstriction, and increases pulmonary vascular resistance; thus, it remains an interesting therapeutic target for CTD-related RP.[49,50] Local therapies such as digital Botox injections may prevent the oxidant-induced intracellular accumulation of reactive oxygen species in vascular endothelial cells and may be considered an important pain management strategy.[51,52] The identification of neuropathic pain may benefit from adjuvants such as gabapentinoids and antidepressants, with temporary use of opioids sometimes required in severe cases with digital ischemia.[53]

Opportunities for clinical trials for RP

It is critical to study primary and CTD-related RP concurrently to clarify medication effectiveness, but it is important to stratify patients based on capillaroscopy, disease severity, and duration. The FDA defines a basket trial as a master protocol study designed to test a single investigational drug or drug combination in different populations defined by disease stage, histology, number of prior therapies, genetic or other biomarkers, or demographic characteristics. This design allows a strong response signal seen in a substudy (ie, SSc-RP) to allow for expansion to generate further data to support regulatory approval.[54] The use of platform clinical trial design may be possible in unique populations, such as the Veterans Health Administration (VHA). Our previous study data from the VHA suggests that CCB medications are potentially being underused for RP and SSc treatment.[55] Furthermore, well-defined disease populations, such those in the Collaborative National Quality and Efficacy Registry for SSc, allow integration of disease duration into outcome assessments, such as pain related to RP and DU.[56]

Conclusions

Treatment of pain related to RP is a challenge for health care providers in the United States due to diagnostic and training limitations and therapeutic access to guideline-based vasodilator medications, such as cost-prohibitive prostanoids. The use of capillaroscopy can allow the identification of patients with CTD

who require medication therapy. Advances in outcome measures, such as the ASRAP, will allow improved clinical trial design for RP that can meet FDA rigor for significant impact on feel and function. Preclinical models and secondary outcomes, which focus on pericyte biology and mechanism of fibrotic reversal, offer hope to patients with SSc with painful DU. Clinical trial design should include both primary and CTD-related RP to clarify therapeutic effect and the mechanism of painful skin blanching to best understand patterns of pain-related to RP in the United States.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Garner R, Kumari R, Lanyon P, et al. Prevalence, risk factors and associations of primary Raynaud's phenomenon: systematic review and meta-analysis of observational studies. BMJ Open 2015;5(3): e006389.
2. Belch J, Carlizza A, Carpentier PH, et al. ESVM guidelines - the diagnosis and management of Raynaud's phenomenon. Vasa 2017; 46(6):413–423.
3. Smith V, Ickinger C, Hysa E, et al. Nailfold capillaroscopy. Best Pract Res Clin Rheumatol 2023;37(1):101849.
4. Nitsche A. Raynaud, digital ulcers and calcinosis in scleroderma. Reumatol Clin 2012;8(5):270–277.
5. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/-European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65(11):2737–2747.
6. Smith V, Herrick AL, Ingegnoli F, et al; EULAR Study Group on Microcirculation in Rheumatic Diseases and the Scleroderma Clinical Trials Consortium Group on Capillaroscopy. Standardisation of nailfold capillaroscopy for the assessment of patients with Raynaud's phenomenon and systemic sclerosis. Autoimmun Rev 2020;19(3): 102458.
7. Pauling JD, Frech TM. Pauling and Frech reply. J Rheumatol 2019; 46(11):1544–1545.
8. Mo YQ, Yan Q, Ye S, et al. Standardized diagnosis and treatment of undifferentiated connective tissue disease and mixed connective tissue disease. Article in Chinese. Zhonghua Nei Ke Za Zhi 2022; 61(10):1119–1127.
9. Del Galdo F, Lescoat A, Conaghan PG, et al. EULAR recommendations for the treatment of systemic sclerosis: 2023 update. Ann Rheum Dis 2025;84(1):29–40.
10. Fernández-Codina A, Walker KM, Pope JE; Scleroderma Algorithm Group. Treatment algorithms for systemic sclerosis according to experts. Arthritis Rheumatol 2018;70(11):1820–1828.
11. Denton CP, Hughes M, Gak N, et al; BSR and BHPR Standards, Guidelines and Audit Working Group. BSR and BHPR guideline for the treatment of systemic sclerosis. Rheumatology (Oxford) 2016; 55(10):1906–1910.
12. Breen ID, Brumfiel CM, Patel MH, et al. Evaluation of the safety of calcitonin gene-related peptide antagonists for migraine treatment among adults with Raynaud phenomenon. JAMA Netw Open 2021; 4(4):e217934.
13. Nawaz I, Nawaz Y, Nawaz E, et al. Raynaud's phenomenon: reviewing the pathophysiology and management strategies. Cureus 2022; 14(1):e21681.
14. Lledó-Ibáñez GM, Sáez Comet L, Freire Dapena M, et al. CAPI-Detect: machine learning in capillaroscopy reveals new variables influencing diagnosis. Rheumatology (Oxford) 2025;64(6):3667–3675.
15. Gracia Tello BDC, Sáez Comet L, Lledó G, et al. Capi-score: a quantitative algorithm for identifying disease patterns in nailfold videocapillaroscopy. Rheumatology (Oxford) 2024;63(12):3315–3321.
16. Ennis H, Hughes M, Anderson ME, et al. Calcium channel blockers for primary Raynaud's phenomenon. Cochrane Database Syst Rev 2016;(2):CD002069.
17. Rirash F, Tingey PC, Harding SE, et al. Calcium channel blockers for primary and secondary Raynaud's phenomenon. Cochrane Database Syst Rev 2017;(12):CD000467.
18. Pauling JD, Frech TM, Domsic RT, et al. Patient participation in patient-reported outcome instrument development in systemic sclerosis. Clin Exp Rheumatol 2017;35(4 suppl 106):184–192.
19. Furst D, Khanna D, Matucci-Cerinic M, et al. Systemic sclerosis - continuing progress in developing clinical measures of response. J Rheumatol 2007;34(5):1194–1200.
20. Merkel PA, Herlyn K, Martin RW, et al; Scleroderma Clinical Trials Consortium. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. Arthritis Rheum 2002;46(9):2410–2420.
21. Black CM, Halkier-Sørensen L, Belch JJ, et al. Oral iloprost in Raynaud's phenomenon secondary to systemic sclerosis: a multicentre, placebo-controlled, dose-comparison study. Br J Rheumatol 1998; 37(9):952–960.
22. Wigley FM, Korn JH, Csuka ME, et al. Oral iloprost treatment in patients with Raynaud's phenomenon secondary to systemic sclerosis: a multicenter, placebo-controlled, double-blind study. Arthritis Rheum 1998;41(4):670–677.
23. Pauling JD, Smith T, Domsic RT, et al. Treatment efficacy in secondary Raynaud's phenomenon. Lancet Rheumatol 2020;2(3):e132.
24. Pauling JD, Reilly E, Smith T, et al. Evolving symptom characteristics of Raynaud's phenomenon in systemic sclerosis and their association with physician and patient-reported assessments of disease severity. Arthritis Care Res (Hoboken) 2019;71(8):1119–1126.
25. DiRenzo DD, Smith TR, Frech TM, et al. Effect of coping strategies on patient and physician perceptions of disease severity and disability in systemic sclerosis. J Rheumatol 2021;48(10):1569–1573.
26. Pauling JD, Reilly E, Smith T, et al. Factors influencing Raynaud condition score diary outcomes in systemic sclerosis. J Rheumatol 2019; 46(10):1326–1334.
27. Virgili-Gervais G, Matthews B, Nassar EL, et al; Scleroderma Patient-centered Intervention Network investigators. The association of outdoor temperature and self-reported Raynaud's phenomenon severity among people with systemic sclerosis: a scleroderma patient-centered intervention network cohort study. Lancet Rheumatol 2024;6(10):e684–e692.
28. Pauling JD, Yu L, Frech TM, et al. Construct validity and reliability of the Assessment of Systemic Sclerosis-Associated Raynaud's Phenomenon (ASRAP) questionnaire. Rheumatology (Oxford) 2024; 63(5):1281–1290.

29. Yu L, Domsic RT, Saketkoo LA, et al. Assessment of the systemic sclerosis-associated Raynaud's phenomenon questionnaire: item bank and short-form development. Arthritis Care Res (Hoboken) 2023;75(8):1725–1734.
30. Scolnik M, Vasta B, Hart DJ, et al. Symptoms of Raynaud's phenomenon (RP) in fibromyalgia syndrome are similar to those reported in primary RP despite differences in objective assessment of digital microvascular function and morphology. Rheumatol Int 2016;36(10):1371–1377.
31. Herrick AL, Murray A. The role of capillaroscopy and thermography in the assessment and management of Raynaud's phenomenon. Autoimmun Rev 2018;17(5):465–472.
32. Maverakis E, Patel F, Kronenberg DG, et al. International consensus criteria for the diagnosis of Raynaud's phenomenon. J Autoimmun 2014;48–49:60–65.
33. Maciejewska M, Sikora M, Maciejewski C, et al. Raynaud's phenomenon with focus on systemic sclerosis. J Clin Med 2022;11(9):2490.
34. Denton CP, De Lorenzis E, Roblin E, et al. Management of systemic sclerosis: British Society for Rheumatology guideline scope. Rheumatol Adv Pract 2023;7(1):rkad022.
35. Flavahan NA. A vascular mechanistic approach to understanding Raynaud phenomenon. Nat Rev Rheumatol 2015;11(3):146–158.
36. Rius Rigau A, Li YN, Matei AE, et al. Characterization of vascular niche in systemic sclerosis by spatial proteomics. Circ Res 2024;134(7):875–891.
37. Elfont RM, Sundaresan PR, Sladek CD. Adrenergic receptors on cerebral microvessels: pericyte contribution. Am J Physiol 1989;256(1 Pt 2):R224–R230.
38. Mourad JJ, Priollet P, Girerd X, et al. The wall to lumen ratio of the radial artery in patients with Raynaud's phenomenon. J Vasc Res 1997;34(4):298–305.
39. Frech T, Walker AE, Barrett-O'Keefe Z, et al. Systemic sclerosis induces pronounced peripheral vascular dysfunction characterized by blunted peripheral vasoreactivity and endothelial dysfunction. Clin Rheumatol 2015;34(5):905–913.
40. Machin DR, Clifton HL, Richardson RS, et al. Acute oral tetrahydrobiopterin administration ameliorates endothelial dysfunction in systemic sclerosis. Clin Exp Rheumatol 2017;35(4 suppl 106):167–172.
41. Machin DR, Clifton HL, Wray DW, et al. Tetrahydrobiopterin administration augments exercise-induced hyperemia and endothelial function in patients with systemic sclerosis. Front Med (Lausanne) 2022;8:791689.
42. Choi YK, Kim YM. Regulation of endothelial and vascular functions by carbon monoxide *via* crosstalk with nitric oxide. Front Cardiovasc Med 2021;8:649630.
43. Mehta KK, Tiwaskar M, Kasture P. Cilnidipine, a dual L/N-type Ca^{2+} channel blocker in hypertension management: a review. J Assoc Physicians India 2024;72(4):54–58.
44. Sternlicht A, Todd M. AISA 021, a novel calcium channel antagonist in development for Raynaud's & systemic sclerosis, has antagonistic activity at sodium channel targets for pain relief and treats scleroderma pain better than current calcium channel blockers in a phase 2A study [abstract]. Arthritis Rheumatol 2024;76(suppl 9). https://acrabstracts.org/abstract/aisa-021-a-novel-calcium-channel-antagonist-in-development-for-raynauds-systemic-sclerosis-has-antagonistic-activity-at-sodium-channel-targets-for-pain-relief-and-treats-scleroderma-pain-better/
45. Fernández-Codina A, Cañas-Ruano E, Pope JE. Management of Raynaud's phenomenon in systemic sclerosis-a practical approach. J Scleroderma Relat Disord 2019;4(2):102–110.
46. Kim HD, Kim EN, Lim JH, et al. Phosphodiesterase inhibitor ameliorates senescent changes of renal interstitial pericytes in aging kidney. Aging Cell 2024;23(3):e14075.
47. Du B, Luo M, Ren C, et al. PDE4 inhibitors for disease therapy: advances and future perspective. Future Med Chem 2023;15(13):1185–1207.
48. Fan T, Wang W, Wang Y, et al. PDE4 inhibitors: potential protective effects in inflammation and vascular diseases. Front Pharmacol 2024;15:1407871.
49. Sagonas I, Daoussis D. Serotonin and systemic sclerosis. An emerging player in pathogenesis. Joint Bone Spine 2022;89(3):105309.
50. Dees C, Chakraborty D, Distler JHW. Cellular and molecular mechanisms in fibrosis. Exp Dermatol 2021;30(1):121–131.
51. Lawson O, Sisti A, Konofaos P. The use of botulinum toxin in Raynaud phenomenon: a comprehensive literature review. Ann Plast Surg 2023;91(1):159–186.
52. Uchiyama A, Yamada K, Perera B, et al. Protective effect of botulinum toxin A after cutaneous ischemia-reperfusion injury. Sci Rep 2015;5(1):9072.
53. van der Gaag A, Cohen SP, Stojanovic MP, et al. 12. Vascular pain: ischemic pain in the extremities and Raynaud's syndrome. Pain Pract 2024;25(1):e13421.
54. Hobbs BP, Pestana RC, Zabor EC, et al. Basket trials: review of current practice and innovations for future trials. J Clin Oncol 2022;40(30):3520–3528.
55. Frech TM, Murtaugh MA, Amuan M, et al. The frequency of Raynaud's phenomenon, very early diagnosis of systemic sclerosis, and systemic sclerosis in a large Veteran Health Administration database. BMC Rheumatol 2021;5(1):42.
56. Shanmugam VK, Frech TM, Steen VD, et al. Collaborative National Quality and Efficacy Registry (CONQUER) for scleroderma: outcomes from a multicenter US-based systemic sclerosis registry. Clin Rheumatol 2020;39(1):93–102.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 113–122
DOI 10.1002/acr.25626

American College of Rheumatology
Empowering Rheumatology Professionals

Longitudinal Trajectories of Pain Sensitization Over Nine Years in People With or at Risk of Knee Osteoarthritis

Kosaku Aoyagi,[1] Na Wang,[2] Michael LaValley,[3] Sarah Tilley,[2] Laura Frey-Law,[4] Cora E. Lewis,[5] Michael Nevitt,[6] and Tuhina Neogi[2]

Objective. Pain sensitization is common in knee osteoarthritis (OA) and is associated with pain severity and functional limitations. Whether pain sensitization is induced by OA or it may be an inherent trait remains unclear. We evaluated pain sensitization trajectories and their relations to symptoms in people with or at risk of knee OA.

Methods. We used data from four study visits of the Multicenter Osteoarthritis Study over a nine-year period. We agnostically identified pain sensitization trajectory groups defined by wrist and knee pressure pain thresholds (PPTs) over nine years using group-based trajectory methods stratified by sex. We evaluated the relations of the wrist and knee PPT trajectory groups to Western Ontario and McMaster Universities Arthritis Index (WOMAC) knee pain and function at the final visit using separate linear regression models.

Results. We included 1,271 participants (mean age 65 years, 61% female, mean body mass index 31 kg/m^2) and identified three distinct trajectory groups in both women and men: low, moderate, and high PPTs. Overall, the PPT trajectories changed minimally over time in each group. Compared to the low PPT trajectory groups (the most sensitized groups), the moderate and high PPT groups had better WOMAC pain and function at the final assessment nine years later.

Conclusion. We identified stable but distinct pain sensitization trajectories over nine years despite likely changes in disease course and treatments over time, suggesting that individuals with knee OA may be predisposed to having different degrees of sensitization as an inherent trait, which in turn likely influences their pain experience and functional status.

INTRODUCTION

Osteoarthritis (OA) commonly affects the knee and has a worldwide prevalence of ~600 million people.[1,2] Knee OA is generally a progressive disease, albeit slowly, over time, and there are no current treatments available to halt or alter its progression.[3] Knee pain is the primary symptom of knee OA and is a leading cause of functional limitations.[1,4,5] There is a recognized discordance between pain and structural (radiographic) pathology in knee OA, suggesting that other factors must be importantly influencing the pain experience.[3] One such factor that is increasingly recognized to contribute to pain severity in OA is pain sensitization, which refers to altered nociceptive signaling in the peripheral and/or central nervous system, with nociceptors becoming more responsive to subthreshold input and even non-noxious input.[3] In addition to contributing to pain severity, measures of pain sensitization are also associated with functional limitations.[6–8] A key question remains as to what causes pain sensitization. Understanding the underpinnings of pain sensitization would provide additional avenues for more comprehensive management of symptomatic OA, a disease for which there are limited efficacious and safe treatment options.

Animal models have demonstrated that pain sensitization can be induced through increased peripheral nociceptive input from chronic inflammation or joint injury.[9,10] If this were the case in human OA, then one would expect that pain sensitization would

Funding sources had no role in study design, data collection and analysis, data interpretation, or the decision to submit the manuscript for publication.

The Multicenter Osteoarthritis Study was supported by the NIH (grants U01-AG-18820, U01-AG-18832, U01-AG-18947, U01-AG-19069, U19-AG-076471, and AR-47785). Dr Neogi's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grants P30-AR-072571, K24-AR-070892, and R01-AG-066010).

[1]University of Texas at El Paso; [2]Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts; [3]Boston University, Boston, Massachusetts; [4]University of Iowa, Iowa City; [5]University of Alabama at Birmingham; [6]University of California, San Francisco.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25626).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25626.

Address correspondence via email to Tuhina Neogi, MD, PhD, at tneogi@bu.edu.

Submitted for publication February 9, 2025; accepted in revised form July 7, 2025.

SIGNIFICANCE & INNOVATIONS

- Pain severity and functional limitations in people with knee osteoarthritis (OA) are influenced by pain sensitization, but whether pain sensitization changes over time in response to OA worsening or remains stable as an inherent trait is not known.
- We found that pain sensitization trajectories remained stable over nine years, despite likely progression of OA, suggesting pain sensitization as an inherent trait rather than a consequence of OA. Further, participants likely used various treatments over the nine-year period, indicating that existing treatments do not meaningfully impact pain sensitization.
- These findings underscore the need for novel treatment strategies targeting pain sensitization as a means to improve pain and function in knee OA.

worsen as OA develops and progresses due to increased peripheral nociceptive input from OA pathology. Conversely, if OA pathology contributes substantially to pain sensitization, one might also expect that interventions such as total knee replacement could lead to improvements in sensitization. Additionally, psychological symptoms and emotional distress are associated with pain sensitization even without actual tissue damage in the periphery.[11] Centrally acting pain medications such as duloxetine, which is approved for management of OA, may also be used for symptomatic relief and could potentially affect pain sensitization.[12] Thus, it is possible that patterns of pain sensitization may change or fluctuate over time because such associated factors or treatments may change over time (ie, pain sensitization as a state, a temporary condition influenced by modifiable factors). On the other hand, there are some recent data suggesting that pain sensitization may be an inherent trait that is present regardless of the severity of radiographic features of OA and may predate development of persistent knee pain.[6,7] With this latter scenario, one would expect that pain sensitization measures would remain relatively stable over time regardless of the course of OA pathology and associated psychological comorbidities or treatments (ie, pain sensitization as an inherent trait, a stable predisposition that varies across individuals).

A greater understanding of pain sensitization patterns and their impact over time would provide important insights into whether pain sensitization may be an inherent trait that is not influenced meaningfully by OA pathology. Such insights would, in turn, provide greater impetus for developing treatment strategies targeting pain sensitization as an additional avenue for managing the symptom burden of OA and clinical assessments of pain sensitization to aid symptomatic prognosis in OA. Thus, we sought to evaluate patterns of pain sensitization over time and the relation of those patterns to pain severity and function using data from a large cohort of people with or at risk of knee OA.

MATERIALS AND METHODS

Study sample. We used data from the Multicenter Osteoarthritis Study (MOST). The MOST is a National Institutes of Health–funded longitudinal cohort study of 3,026 older adults aged between 50 and 79 years old with or at risk of knee OA. Knee OA status was determined by radiographic evaluation using Kellgren–Lawrence grade ≥2. Risk of knee OA was defined by the presence of knee symptoms, being overweight, or a history of knee injury or surgery. Details of the study protocol have been published elsewhere.[13–15] The study protocol was approved by the institutional review boards at the University of Iowa; University of Alabama, Birmingham; University of California, San Francisco; and Boston University School of Medicine. The first study visit at which measures of pain sensitization were obtained was at the fifth-year visit. Thus, the fifth-year visit served as the baseline for these analyses. For these analyses, we included data from the subsequent study visits that occurred over a total 9-year period: 7th-, 12th-, and 14th-year study visits, reflecting 2, 7, and 9 years of follow-up, respectively. We excluded those with peripheral neuropathy, rheumatoid arthritis, and other inflammatory arthritis from the study. We also excluded data post–knee replacement surgery as needed (see statistical analysis section for details) because such surgeries could potentially alter pain sensitivity due to the procedure itself and thereby limit our ability to evaluate the natural longitudinal trajectories of pain sensitization over time.

Measures of pain sensitization. *Pressure pain threshold.* We evaluated pressure pain threshold (PPT) as the measure of pain sensitization for these analyses. PPT is a reliable measure of pain sensitization evoked by mechanical stimulation.[6,16] PPT at the primary site of pathology (eg, knee) is thought to reflect peripheral sensitization with or without central sensitization, whereas PPT at a distant body site unrelated to the primary site of pathology, such as the wrist, is thought to reflect central sensitization because it reflects sensitivity outside the area of pain.[6,17] We measured PPT at the right distal radioulnar joint (wrist) and the patella (knee) using a handheld pressure algometer (1-cm^2 rubber tip; Wagner FDIX25) applied at a constant rate of 0.5 kg/second, with values recorded in kilograms per square centimeter. PPT was defined as the point at which the participant indicated that the pressure first changed to slight pain. The PPT at each anatomic site was calculated by averaging three trials. Lower PPT values indicate greater pain sensitivity. PPTs were performed by trained staff, with demonstrated accuracy relative to a master examiner for staff to be certified. Fourteen-day test–retest reliability of PPT within the MOST was excellent, with intraclass correlation coefficients ranging from 0.85 to 0.90, indicating high consistency of repeated measurements.[6]

Clinical outcomes: pain severity and function. The Western Ontario and McMaster Universities Arthritis Index (WOMAC) pain subscale (0–20 score) and function subscale (0–68) were used to assess pain severity and physical function at the final study visit at the end of nine years of follow-up.[18] The pain and function subscales consist of 5 questions and 17 questions, respectively, and each is assessed on a 0 to 4 Likert scale (none [0], mild [1], moderate [2], severe [3], and extreme [4]). The WOMAC pain subscale was obtained as a knee-based assessment, whereas the WOMAC function was obtained as a person-based assessment. Higher WOMAC scores represent greater pain severity and worse physical function.

Statistical analysis. To assess patterns of pain sensitization over time, we used discrete group-based trajectory methods using the SAS macro PROC TRAJ procedure to agnostically identify pain sensitization trajectory groups using wrist and knee PPT data over the four time points across nine years. This group-based trajectory method is a data-driven approach that agnostically identifies patterns or shapes of change over time without assumptions, distributes individual observations to the corresponding trajectory group, and estimates missing data using maximum likelihood estimation[19–23] under the standard assumption of data missing at random. To determine the patterns of pain sensitization trajectories, we included both linear and quadratic terms to capture linear and/or curvilinear trends.[21]

For knee PPT, we used the more painful knee or selected the right knee if both knees were equally painful. We included PPTs from those who had knee replacement surgery if they had a minimum of two PPT time points available before knee replacement because PROC TRAJ accommodates missing data.[19–24] These participants could still be included in trajectory estimation based on their available presurgery data. We stratified trajectory group assessments by sex because prior studies have consistently shown lower PPT values in women patients than in men patients, which may reflect sex-related differences in pain processing related to pain sensitization.[25–27] Therefore, four different sets of pain trajectory analyses were performed (ie, wrist PPT trajectories in women and men and knee PPT trajectories in women and men).

We required a trajectory group to have a minimum of 5% of the sample to avoid overfitting and provide clinically meaningful and distinct patterns in pain sensitization over time.[21] We assessed model fit using Akaike information criterion, Bayesian information criterion, and posterior probability of group membership,[20,21,23] which evaluates the likelihood of individual observations belonging to each trajectory group. A higher posterior probability (ie, closer to 1.0) indicates a higher likelihood an individual's trajectory pattern fits within the border of the trajectory group. A posterior probability of at least 0.7 is recommended as a minimum threshold for adequate model fit.[19] We assigned each observation to the trajectory group with the highest posterior probability of group membership. We compared multiple model specifications (eg, three- to six-group solutions) and selected the three-group model based on the combination of sample size, model fit statistics, and clinical interpretability. For each trajectory group, we described their general characteristics, including mean (±SD) age, percentage female, mean (±SD) body mass index (BMI), percentage White, percentage with depressive symptoms (based on cutoff of ≥16 on Center for Epidemiologic Studies Depression Scale[28]), percentage with catastrophizing (one item from the Coping Strategies Questionnaire, with score ≥1 defined as the presence of mild or greater pain catastrophizing[29]), and percentage with radiographic knee OA, defined as presence of tibiofemoral OA (Kellgren–Lawrence grade ≥ 2) and/or patellofemoral OA (osteophyte score ≥2, or joint space narrowing score ≥2 with any osteophyte, and sclerosis or cyst score ≥1 on the lateral view), or knee replacement.

We then evaluated the relations of the trajectory groups to WOMAC knee pain at the final study visit (ie, fourth PPT assessment visit) using knee-based linear regression models with generalized estimating equations to account for the correlation between two knees within an individual for WOMAC knee pain. Similarly, we evaluated relations of the PPT trajectory groups to WOMAC function using separate person-based linear regression models. In the main analyses, we performed Tobit regression models because our outcome values are negatively skewed, with >25% of the sample having a value of 0 for WOMAC pain or function.[30,31]

In addition, we performed sensitivity analyses limited to those with a patient acceptable symptom state (PASS) for pain (ie, WOMAC pain ≥ 5.35/20)[32] and for function (ie, WOMAC function ≥ 21.08/68),[33] respectively, at baseline. This was done in an attempt to minimize the potential for reverse causation because those with the worst pain at baseline may also be the most likely to have the greatest sensitization, and thus, it would be expected that this group would continue to have the worst pain at the final study visit. We used the same discrete group-based trajectory methods in these samples as described above. We then evaluated the relations of the PPT trajectory groups to the risk of failing to maintain PASS at the final study visit using separate log-binomial regression models. All analyses were adjusted for the following potential confounders: age, sex (female or male), BMI (a continuous variable), race and ethnicity (a nominal variable as a proxy for numerous differential structural systemic societal impacts), clinic site (data collection sites of the MOST in either Iowa or Alabama), depressive symptoms (a binary variable), and pain catastrophizing (a binary variable). All analyses were conducted using SAS version 9.4 (SAS Institute).

Data availability and ethics approval. Deidentified data were used for this work. Investigators may request data access from the MOST investigators. This study received approval from the following: Boston University Medical Campus Institutional Review Board (protocol H-32821), The University of

Iowa Institutional Review Board (201511711), University of Alabama at Birmingham Institutional Review Board (000329007), and University of California, San Francisco, Institutional Review Board (10-00500).

RESULTS

A total of 1,271 participants were included in this study (mean age 65 years, 61% female, 88% White, and mean BMI 31 kg/m^2). Radiographic knee OA, depressive symptoms, and pain catastrophizing were present in 60%, 8%, and 53% of the participants, respectively.

PPT trajectory groups. We identified distinct wrist and knee PPT trajectory groups over nine years. Each trajectory group was linear; we did not find any statistically significant quadratic terms indicating the presence of curvilinear trajectory patterns.

Wrist PPT trajectory groups. We identified three distinct wrist PPT trajectory groups over nine years for both women and men, which we have characterized as high, moderate, and low PPT groups, with the lowest PPTs indicating the most pain-sensitized groups (Figure 1). Low, moderate, and high wrist PPT trajectory groups had a prevalence of 71.7%, 22.4%, and 5.9% among women and 53.9%, 36.1%, and 10.0% among men, respectively. The mean posterior probability ranged from 0.84 to 0.98 for women and 0.75 to 0.98 for men, respectively, indicating good to excellent probabilities that each observation belongs to the corresponding trajectory groups. The trajectories of the moderate and low PPT groups for both women and men remained relatively flat over nine years. The trajectories of the high PPT groups for women and men changed minimally, with women having decreased slightly, whereas men increased slightly over the nine years, though all changes were within a 1-kgf/cm^2 range, which is less than a meaningful difference in knee OA.[34,35]

Baseline characteristics of the wrist PPT trajectory groups are presented in Table 1. Demographics and the presence of radiographic knee were similar across all trajectory groups: mean age ~65 years, ~83% White, mean BMI ~30 kg/m^2, and ~55% with radiographic knee OA. The low and moderate wrist PPT groups (ie, the more sensitized groups) had a higher percentage of women with depressive symptoms than the high PPT group (12%, 11%, and 7%, respectively), whereas low, moderate, and high wrist PPT trajectory groups had 5%, 3%, and 10% of men who had depressive symptoms, respectively. The low PPT groups had the highest proportion of people with catastrophizing for both women and men.

Knee PPT trajectory groups. Similar to the wrist PPT trajectory groups, we identified three knee PPT trajectory groups for both women and men separately, which we also characterized as low, moderate, and high PPT groups (low PPT group indicates the most

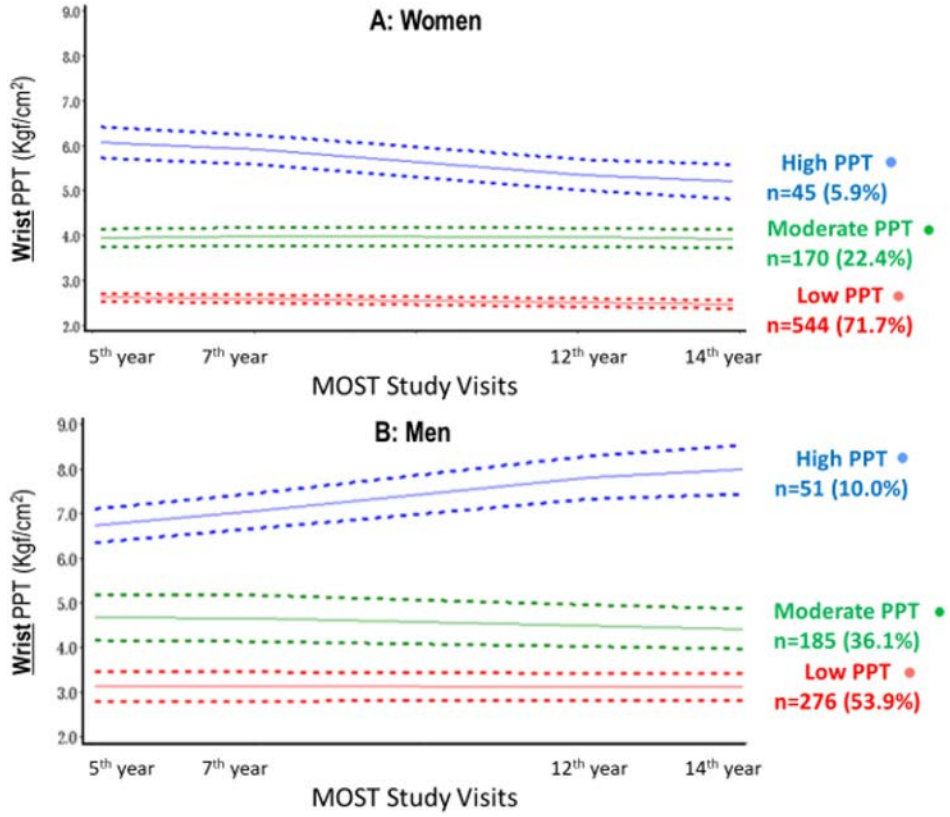


Figure 1. Wrist PPT trajectory groups: (A) women and (B) men. Solid lines are the estimates for the wrist PPT trajectories over nine years, and dashed lines are 95% confidence intervals around the estimates. MOST, Multicenter Osteoarthritis Study; PPT, pressure pain threshold.

Table 1. Participant characteristics of wrist PPT trajectory groups*

	Women			Men		
Wrist PPT trajectory groups	Low (n = 544)	Moderate (n = 170)	High (n = 45)	Low (n = 276)	Moderate (n = 185)	High (n = 51)
Age, mean (±SD), y	65.6 (±7.2)	65.4 (±7.0)	63.1 (±6.7)	64.9 (±7.3)	65.1 (±7.2)	62.5 (±6.3)
White, n (%)	435 (80)	141 (83)	39 (86)	231 (84)	163 (88)	46 (90)
BMI, mean (±SD), kg/m^2	30.1 (±6.0)	30.8 (±6.6)	32.4 (±8.5)	30.3 (±5.2)	30.2 (±4.9)	31.0 (±4.5)
Depressive symptom, n (%)	63 (11.6)	18 (10.6)	3 (6.7)	14 (5.1)	5 (2.7)	5 (9.8)
Catastrophizing, n (%)	288 (52.9)	72 (42.4)	17 (37.8)	124 (44.9)	58 (31.4)	12 (23.5)
Radiographic knee OA, n (%)	293 (53.8)	100 (58.8)	28 (62.2)	160 (58.0)	93 (50.3)	26 (51.0)
Knee OA KL grade, n (%)						
0	194 (35.7)	56 (32.9)	15 (13.3)	105 (38.0)	84 (45.4)	32 (45.1)
1	63 (11.6)	19 (11.2)	5 (11.1)	32 (11.6)	30 (16.2)	6 (11.8)
2	122 (22.4)	41 (24.1)	10 (22.2)	51 (18.5)	20 (10.8)	4 (7.8)
3	113 (20.8)	35 (20.6)	8 (17.8)	44 (15.9)	37 (20.0)	11 (21.6)
4	38 (7.0)	16 (9.4)	6 (13.3)	33 (12.0)	10 (5.4)	5 (9.8)

* Radiographic knee OA was defined as the presence of tibiofemoral OA (KL grade ≥ 2) and/or patellofemoral OA (osteophyte score ≥2 or joint space narrowing score ≥2 with any osteophyte, sclerosis, or cyst score ≥1 on the lateral view). BMI, body mass index; KL, Kellgren–Lawrence; OA, osteoarthritis; PPT, pressure pain threshold.

pain-sensitized group; Figure 2). Low, moderate, and high knee PPT trajectory groups had a prevalence of 55.8%, 32.5%, and 11.7% among women and a prevalence of 21.1%, 44.5%, and 34.4% among men, respectively. The median posterior probability ranged from 0.83 to 0.97 for women and 0.84 to 0.97 for men, respectively, indicating good to excellent probabilities that each observation belongs to the corresponding trajectory groups. The PPTs in each of the trajectory groups remained fairly stable over time. Baseline characteristics of the knee PPT trajectory groups were similar to the wrist PPT trajectory groups, with

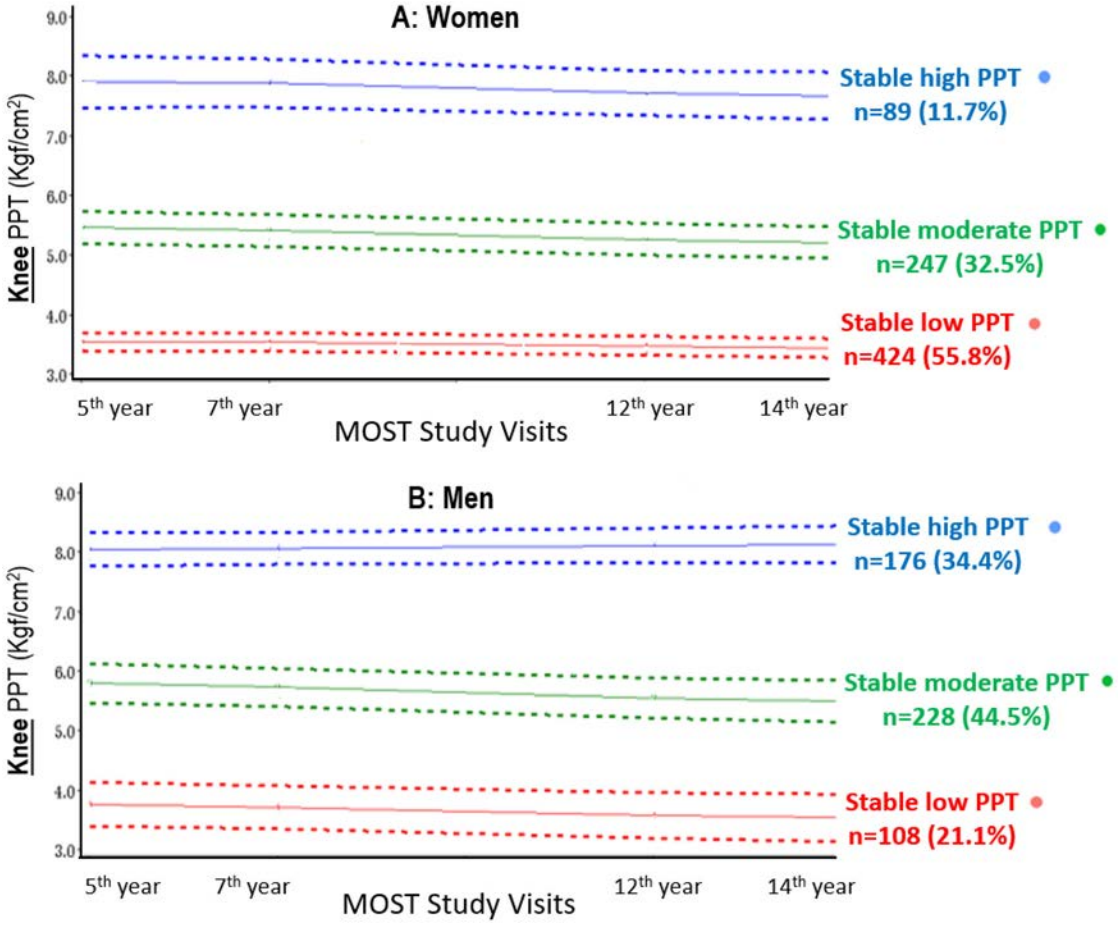


Figure 2. Knee PPT trajectory groups: (A) women and (B) men. Solid lines are estimates for the wrist PPT trajectories over nine years. Dashed lines are 95% confidence intervals around the estimates. MOST, Multicenter Osteoarthritis Study; PPT, pressure pain threshold.

Table 2. Participant characteristics of knee PPT trajectory groups*

	Women			Men		
Knee PPT trajectory groups	Low (n = 424)	Moderate (n = 247)	High (n = 89)	Low (n = 108)	Moderate (n = 228)	High (n = 176)
Age, mean (±SD), y	65.4 (±7.2)	65.6 (±7.2)	64.8 (±6.6)	64.0 (±7.4)	64.9 (±7.4)	65.0 (±6.9)
White, n (%)	381 (90)	204 (83)	76 (86)	99 (92)	210 (92)	141 (80)
BMI, mean (±SD), kg/m^2	31.3 (±6.5)	29.4 (±6.0)	29.0 (±5.5)	31.0 (±5.3)	30.5 (±5.2)	29.8 (±4.5)
Depressive symptom, n (%)	54 (12.7)	21 (7.95)	9 (10.1)	9 (8.0)	9 (4.0)	6 (3.4)
Catastrophizing, n (%)	220 (51.9)	124 (47.0)	33 (37.8)	48 (42.5)	90 (40.4)	56 (31.8)
Radiographic knee OA, n (%)	224 (52.8)	146 (56.4)	51 (58.0)	61 (57.0)	127 (59.4)	91 (52.3)
Knee OA KL grade, n (%)						
0	153 (36.1)	82 (33.3)	30 (33.7)	49 (45.4)	87 (38.2)	76 (43.2)
1	44 (10.4)	35 (14.2)	8 (9.0)	10 (9.3)	35 (15.4)	23 (13.1)
2	93 (21.9)	55 (22.4)	24 (28.1)	22 (20.4)	32 (14.0)	21 (11.9)
3	94 (22.2)	44 (17.9)	18 (20.2)	12 (11.1)	39 (17.1)	41 (23.3)
4	29 (6.8)	24 (9.8)	7 (7.9)	12 (11.1)	24 (10.5)	12 (6.8)

* Radiographic knee OA was defined as the presence of tibiofemoral OA (KL grade ≥ 2) and/or patellofemoral OA (osteophyte score ≥2 or joint space narrowing score ≥2 with any osteophyte, sclerosis, or cyst score ≥1 on the lateral view). BMI, body mass index; KL, Kellgren–Lawrence; OA, osteoarthritis; PPT, pressure pain threshold.

demographic and radiographic knee OA characteristics being similar across trajectory groups (Table 2). The low knee PPT group (ie, most sensitized group) had the highest percentage of people with depressive symptoms and pain catastrophizing.

Relation of pain sensitization trajectories to WOMAC pain and function at the 14th-year study visit. Overall, as compared to the low trajectory group (the most sensitized trajectory group), moderate and high trajectory groups had better WOMAC pain and function scores at the final assessment (ie, after nine years; Figure 3; Supplementary Table 1). In general, both women and men in the moderate and high trajectory groups had WOMAC pain scores that were one to two units lower than those in the low trajectory groups. Similar patterns were seen for WOMAC function, as well as for the knee PPT trajectory groups (Figure 4; Supplementary Table 2).

Pain sensitization trajectories and PASS status at the 14th-year study visit. In the sensitivity analyses, we had 545 participants who met WOMAC pain PASS and

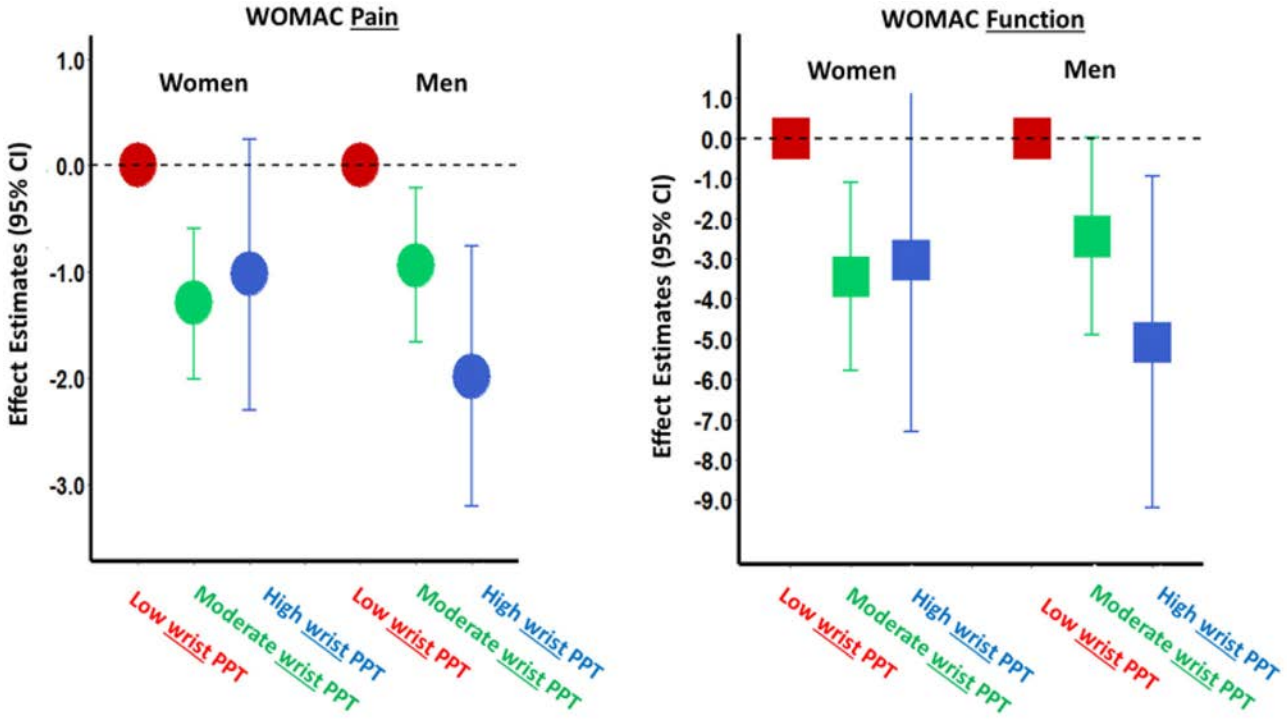


Figure 3. Relation of wrist PPT trajectory groups to WOMAC pain and function at the 14th visit. Black dashed horizontal line denotes the null effect ($\beta = 0$). Estimates whose 95% CIs cross this line are not statistically different from the reference (Low wrist PPT) group; values below the line indicate lower (better) WOMAC pain/function scores. CI, confidence interval; PPT, pressure pain threshold; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

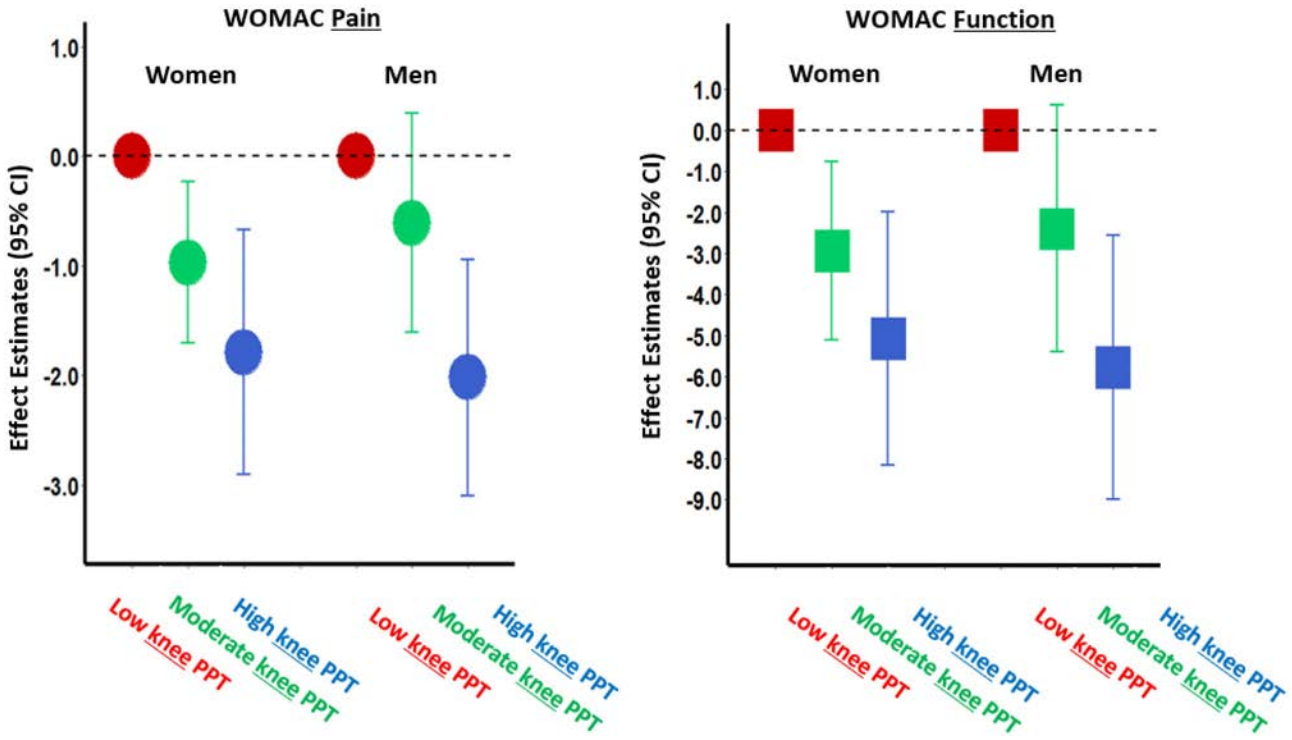


Figure 4. Relation of knee PPT trajectory groups to WOMAC pain and function at the 14th visit. Black dashed horizontal line denotes the null effect ($\beta = 0$). Estimates whose 95% CIs cross this line are not statistically different from the reference (Low wrist PPT) group; values below the line indicate lower (better) WOMAC pain/function scores. CI, confidence interval; PPT, pressure pain threshold; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

594 participants who met WOMAC function PASS at baseline. Similar to the main analyses, we identified three distinct trajectories with moderate and low PPT trajectories being relatively stable over nine years, with an exception for wrist PPT trajectories in women in whom the WOMAC function sample had only two trajectories. The trajectories and corresponding posterior probabilities are summarized in Supplementary Figure 1. Baseline characteristics of PPT trajectory groups among those who met PASS at baseline were similar to the main cohort (summarized in Supplementary Table 3).

Approximately 85% to 90% of the sample within each trajectory group maintained PASS status, whereas ~10% to 15% failed to maintain PASS status at the final assessment at the ninth year of follow-up. To evaluate these trajectory groups and PASS status at the final visit, we combined moderate and high PPT trajectory groups due to the limited sample size. We evaluated the risk of failing to maintain the PASS status at the final visit among the lowest PPT groups (the most sensitized groups) compared with the combined higher PPT groups. Overall, there were no significant associations. The numerical trends indicated that the low PPT groups had a higher risk of failure to maintain the WOMAC pain PASS status at the final visit, but no such trends were noted for WOMAC function PASS status. For example, the low wrist PPT group in women had a 54% greater risk of failing to maintain the WOMAC pain PASS status as compared to the moderate and high wrist PPT groups, but this did not reach statistical significance (adjusted risk ratio 1.54, 95% confidence interval 0.91–2.62; Supplementary Table 4).

DISCUSSION

We found three distinct pain sensitization trajectory groups assessed with either wrist PPT or knee PPT over nine years in both women and men with or at risk of knee OA. The slopes of the trajectories remained relatively stable over nine years, indicating that the degree of sensitization did not appear to worsen or improve over time despite likely changes in OA status and potential therapies used over nine years. Additionally, the trajectory groups were associated with both pain severity and physical function at the final visit (ie, after nine years of follow-up), with the most sensitized group exhibiting the greatest pain severity and functional limitations. These findings provide novel insights into the varying degrees of pain symptoms and functional limitations experienced by individuals with OA.

To our knowledge, this is the first study to identify longitudinal trajectories of pain sensitization in knee OA and evaluate their associations with clinical outcomes. A recent systematic review and meta-analysis reported that pain sensitization assessed with quantitative sensory testing (QST), including PPT, may remain stable over time in people with whiplash-associated disorders,[36] although the conclusion was made based on short-term follow-up (ie, between baseline and 3- or 12-month assessments).[37,38] Our study extends these shorter-term studies to provide data

over multiple time points over nine years. We had hypothesized that pain sensitization trajectories would change or fluctuate over time because OA would be expected to progress over nine years, other factors such as psychological symptoms may change over time, and different treatment modalities may be instituted, which in turn may potentially influence pain sensitization.[11] However, in our study, we noted distinct and stable pain sensitization trajectories over nine years. This finding suggests that individuals might be predisposed to have different degrees of pain sensitivity as inherent traits, or at least due to factors unrelated directly to OA pathology. Additionally, the lack of changes in sensitization levels over nine years despite the possibility that individuals used various treatments to manage symptoms suggests that existing management options have limited impact on pain sensitization.

We also found that pain and functional limitations were associated with pain sensitivity trajectory groups. Pain sensitization, particularly central sensitization, may disrupt the integrity of sensorimotor and motor cortices through altered nociceptive signaling.[9,39–41] This may involve disinhibition of the motor cortex, potentially driven by reduced GABAergic activity and/or increased levels of brain-derived neurotrophic factor.[42,43] Such changes can impair motor planning and coordination, leading to inefficient or maladaptive movement strategies that contribute to functional limitations, independent of pain intensity.[8,9,42–44] A recent study reported that pain sensitization assessed with QST was cross-sectionally associated with WOMAC function.[8] Our findings add further insights into longitudinal associations of pain sensitization with physical function in knee OA. Sensitivity analyses conducted to address potential reverse causation by examining those with PASS at baseline were limited by precision but generally supported the main findings of greater pain sensitization at baseline being associated with poor pain outcome over time. In contrast, we did not find that the most sensitized group had a greater risk of failing to meet the PASS for function over the nine-year follow-up period, potentially indicating that a longer follow-up may be needed to observe the findings of the main analyses. Alternatively, individuals may have developed compensatory strategies to maintain physical function despite persistent pain sensitivity. These adaptations could help explain why the most sensitized group did not have a greater risk of failing to meet the PASS for function over the nine-year follow-up period.

The observation of similar longitudinal trajectories for wrist and patellar PPT groups may suggest that different individuals have distinct inherent traits both centrally and peripherally or that a shared central component was acting on the knee PPT as well. Additionally, both wrist and patellar PPT trajectory groups had similar results in their associations with pain severity and physical function, which aligned with our prior work in which both peripheral sensitivity and central sensitivity were associated with knee pain severity.[6] Our results were also similar between women and men, though overall, women had lower PPT values than men, which are thought to reflect one dimension of pain sensitization (Supplementary Tables 5 and 6), as reported from prior studies.[25–27]

Our findings provide some clinical implications. First, people with pain sensitization (ie, greater pain sensitivity) may anticipate experiencing more severe symptoms over time than those with less pain sensitivity. Second, the observed stability in pain sensitivity trajectories over nine years, despite the administration of various treatments during this period, suggests a lack of effective treatments targeting pain sensitization to date. The identification of distinct and stable trajectory groups such as stable low PPT may help clinicians recognize subgroups of patients who are at greater risk for persistent symptoms and guide monitoring or support more personalized care. Our findings underscore the need for novel treatment approaches targeting pain sensitization to mitigate knee OA symptoms over time.

We have several limitations to acknowledge. Although we found three distinct trajectory groups for each set of analyses, variability within the groups may have been obscured (ie, subsets that did indeed have some improvement or worsening over time), but overall, those did not dominate sufficiently to identify separate distinct groups with such patterns. The identified trajectories were the best model fit and support that pain sensitization remained stable over nine years for the majority of individuals. Second, the average knee pain severity and functional limitations in our sample were relatively mild. Therefore, our results may not be fully applicable to people with more severe knee pain and/or severe functional limitations. Third, our analysis did not include QST measures of temporal summation or conditioned pain modulation, which are also thought to assess alterations in nociceptive signaling. The absence of these measures may limit the breadth of our evaluation of pain sensitization mechanisms. Fourth, we did not analyze treatments occurring over the nine-year period in relation to trajectory group membership, which limits our ability to interpret the impact of pain management on the stability of sensitization. Nonetheless, it is reasonable to assume that individuals with knee OA likely received various forms of treatment during this time, and despite that, the trajectories remained relatively stable over that time. Finally, >80% of participants were White, reflecting the underlying cohort in the current analysis; thus, our findings may not be applicable to other demographic groups.

In summary, we identified three distinct and stable pain sensitization trajectory groups over nine years for both women and men with or at risk of knee OA, and the different sensitization groups had different levels of knee pain severity and functional limitations. The slopes of the trajectory groups remained relatively stable despite the likelihood of structural OA progression over this time, which may be expected to result in greater nociceptive input or the use of various treatments. Thus, our results suggest that individuals with or at risk of knee OA may be predisposed to having different degrees of pain sensitization as an inherent trait, and this finding may, in part, explain why some individuals experience more or less pain despite similar degrees of OA pathology. These

findings highlight the importance of pain sensitization to symptomatic prognosis in knee OA and underscore a need for developing novel treatment strategies to address pain sensitization as a means to improve pain and function in knee OA.

ACKNOWLEDGMENT

The authors would like to thank all the investigators, research staff, and study participants.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Safiri S, Kolahi AA, Smith E, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020;79(6):819–828.
2. Steinmetz JD, Culbreth GT, Haile LM, et al. Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5(9):e508–e522.
3. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21(9):1145–1153.
4. Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and beyond: a Lancet Commission. Lancet 2020;396(10264):1711–1712.
5. Cross M, Smith E, Hoy D, et al. The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 Study. Ann Rheum Dis 2014;73(7):1323–1330.
6. Neogi T, Frey-Law L, Scholz J, et al. Sensitivity and sensitisation in relation to pain severity in knee osteoarthritis: trait or state? Ann Rheum Dis 2015;74(4):682–688.
7. Carlesso LC, Segal NA, Frey-Law L, et al. Pain susceptibility phenotypes in those free of knee pain with or at risk of knee osteoarthritis: the Multicenter Osteoarthritis Study. Arthritis Rheumatol 2019;71(4):542–549.
8. Corrigan P, Neogi T, Frey-Law L, et al. Relation of pain sensitization to self-reported and performance-based measures of physical functioning: the Multicenter Osteoarthritis (MOST) study. Osteoarthritis Cartilage 2023;31(7):966–975.
9. Latremoliere A, Woolf CJ. Central sensitization: a generator of pain hypersensitivity by central neural plasticity. J Pain 2009;10(9):895–926.
10. Woolf CJ. Central sensitization: uncovering the relation between pain and plasticity. Anesthesiology 2007;106(4):864–867.
11. Meints SM, Edwards RR. Evaluating psychosocial contributions to chronic pain outcomes. Prog Neuropsychopharmacol Biol Psychiatry 2018;87:168–182.
12. Chen B, Duan J, Wen S, et al. An updated systematic review and meta-analysis of duloxetine for knee osteoarthritis pain. Clin J Pain 2021;37(11):852–862.
13. Felson DT, Niu J, Guermazi A, et al. Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. Arthritis Rheum 2007;56(9):2986–2992.
14. Englund M, Felson DT, Guermazi A, et al. Risk factors for medial meniscal pathology on knee MRI in older US adults: a multicentre prospective cohort study. Ann Rheum Dis 2011;70(10):1733–1739.
15. Segal NA, Nevitt MC, Gross KD, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. Pm R 2013;5(8):647–654.
16. Wessel J. The reliability and vatidity of pain threshold measurements in osteoarthritis of the knee. Scand J Rheumatol 1995;24(4):238–242.
17. Neogi T, Guermazi A, Roemer F, et al. Association of joint inflammation with pain sensitization in knee osteoarthritis: the Multicenter Osteoarthritis Study. Arthritis Rheumatol 2016;68(3):654–661.
18. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15(12):1833–1840.
19. Nagin D. Group-Based Modeling of Development. Harvard University Press; 2005.
20. Collins JE, Katz JN, Dervan EE, Losina E. Trajectories and risk profiles of pain in persons with radiographic, symptomatic knee osteoarthritis: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2014;22(5):622–630.
21. White DK, Neogi T, King WC, et al. Can change in prolonged walking be inferred from a short test of gait speed among older adults who are initially well-functioning? Phys Ther 2014;94(9):1285–1293.
22. Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. Sociol Methods Res 2001;29(3):374–393.
23. Pan F, Tian J, Scott D, et al. Muscle function, quality, and relative mass are associated with knee pain trajectory over 10.7 years. Pain 2022;163(3):518–525.
24. Dodge HH, Shen C, Ganguli M. Application of the pattern-mixture latent trajectory model in an epidemiological study with non-ignorable missingness. J Data Sci JDS 2008;6(2):247–259.
25. Gerum M, Simonin F. Behavioral characterization, potential clinical relevance and mechanisms of latent pain sensitization. Pharmacol Ther 2022;233:108032.
26. Fitzcharles MA, Cohen SP, Clauw DJ, et al. Nociplastic pain: towards an understanding of prevalent pain conditions. Lancet 2021;397(10289):2098–2110.
27. Neziri AY, Curatolo M, Nüesch E, et al. Factor analysis of responses to thermal, electrical, and mechanical painful stimuli supports the importance of multi-modal pain assessment. Pain 2011;152(5):1146–1155.
28. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385–401.
29. Jensen MP, Keefe FJ, Lefebvre JC, et al. One-and two-item measures of pain beliefs and coping strategies. Pain 2003;104(3):453–469.
30. Tobin J. Estimation of relationships for limited dependent variables. Econom J Econom Soc 1958;26:24–36.
31. Aoyagi K, Law LF, Carlesso L, et al. Post-surgical contributors to persistent knee pain following knee replacement: the Multicenter Osteoarthritis Study (MOST). Osteoarthr Cartil Open 2023;5:100335.
32. Georgopoulos V, Smith S, McWilliams DF, et al. Harmonising knee pain patient-reported outcomes: a systematic literature review and meta-analysis of patient acceptable symptom state (PASS) and individual participant data (IPD). Osteoarthritis Cartilage 2023;31(1):83–95.
33. Tubach F, Ravaud P, Baron G, et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis:

the minimal clinically important improvement. Ann Rheum Dis 2005; 64(1):29–33.

34. Burrows NJ, Booth J, Sturnieks DL, et al. Acute resistance exercise and pressure pain sensitivity in knee osteoarthritis: a randomised crossover trial. Osteoarthritis Cartilage 2014;22(3):407–414.
35. Arendt-Nielsen L, Nie H, Laursen MB, et al. Sensitization in patients with painful knee osteoarthritis. Pain 2010;149(3):573–581.
36. Bellosta-López P, Doménech-García V, Ortiz-Lucas M, et al. Longitudinal changes and associations between quantitative sensory testing and psychological factors in whiplash-associated disorders: a systematic review and meta-analyses-based data synthesis. J Pain 2024;25:12–30.
37. Sterling M, Vicenzino B, Souvlis T, et al. Dry-needling and exercise for chronic whiplash-associated disorders: a randomized single-blind placebo-controlled trial. Pain 2015;156(4):635–643.
38. Michaleff ZA, Maher CG, Lin CWC, et al. Comprehensive physiotherapy exercise programme or advice for chronic whiplash (PROMISE): a pragmatic randomised controlled trial. Lancet 2014;384(9938): 133–141.
39. Woolf CJ. Pain: moving from symptom control toward mechanism-specific pharmacologic management. Ann Intern Med 2004;140(6): 441–451.
40. Cheng JC, Anzolin A, Berry M, et al. Dynamic functional brain connectivity underlying temporal summation of pain in fibromyalgia. Arthritis Rheumatol 2022;74:700–710.
41. Cagnie B, Coppieters I, Denecker S, et al. Central sensitization in fibromyalgia? A systematic review on structural and functional brain MRI. Semin Arthritis Rheum 2014;44(1):68–75.
42. Roussel NA, Nijs J, Meeus M, et al. Central sensitization and altered central pain processing in chronic low back pain: fact or myth? Clin J Pain 2013;29(7):625–638.
43. Caumo W, Deitos A, Carvalho S, et al. Motor cortex excitability and BDNF levels in chronic musculoskeletal pain according to structural pathology. Front Hum Neurosci 2016;10:357.
44. Aoyagi K, Sharma NK. Correlation between central sensitization and remote muscle performance in individuals with chronic low back pain. J Manip Physiol Ther 2021;44:14–24.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 123–133
DOI 10.1002/acr.25687

American College of Rheumatology
Empowering Rheumatology Professionals

Centering Patient Voices in Lupus Pain: A Biopsychosocial Analysis of Reddit Narratives Using Large Language Models

Andrew Walker,[1] Jerik Leung,[2] Aishwarya Alagappan,[3] Swati Rajwal,[1] Sahithi Lakamana,[1] Tricia Park,[1] Nathan Le,[3] Anushka Irani,[4] Abeed Sarker,[1,5] Titilola Falasinnu,[3] and Selen Bozkurt[1]

Objective. Patients with chronic illness share their experiences in online communities and generate rich data on pain management. This study applied natural language processing methods, including large language models (LLMs), to Reddit discussions from lupus communities to characterize multidimensional pain experiences framed in the biopsychosocial model.

Methods. We extracted Reddit posts from the r/Lupus and r/LupusSupport subreddits posted from June 9, 2010, through December 31, 2023. Pain-related posts were identified using a clinically informed pain lexicon. Topic modeling was used to identify thematic patterns, which were then compared with structured summaries generated by LLM instructions that were fine-tuned using the biopsychosocial model of pain. Two reviewers conducted content analysis of the LLM-generated summaries, evaluating thematic accuracy and coverage.

Results. Data from Reddit included 31,785 posts from 10,857 authors. We identified common pain complaints, management strategies, and sociocultural, affective, and nociplastic dimensions of pain. Instruction fine-tuned LLMs produced structured summaries with an average thematic accuracy score of 3.1 of 4 (kappa = .09) and content coverage score of 2.9 of 4 (kappa = .38). Sociocultural features presented in 123 posts (33.8%), including peer support and validation (n = 106) and provider interactions or access issues (n = 35). Nociplastic pain presented in 205 posts (56.3%).

Conclusion. Natural language processing methods can be used to extract rich, multidimensional insights into pain experiences from online communities focused on lupus. These approaches highlight the psychological, social, and cultural facets of pain that may be underrepresented in clinical settings, supporting more patient-centered approaches to care in rheumatology.

INTRODUCTION

More than half of patients with systemic lupus erythematosus (SLE) experience chronic pain, which may be directly related to inflammation or less clearly linked to SLE flares or disease activity (eg, migraines, low back pain, fibromyalgia), significantly contributing to disability.[1,2] Despite its prevalence, pain in SLE remains undercharacterized, particularly from the patient's perspective.[2] Bridging this gap requires innovative approaches that extend beyond traditional clinical data sources. One such approach involves leveraging patient-generated content from online health communities. Platforms like Reddit provide a unique window into the perspectives of individuals living with SLE. Many aspects of SLE, including pain triggers, symptom variability, and disease flares, remain elusive to both patients and clinicians. As a result, patients often turn to peer networks to process uncertainty, seek validation, and share coping strategies.[3–5]

Recognizing the value of patient voice, the US Food and Drug Administration has encouraged investigating social media to capture patient experiences and perspectives and monitor adverse events.[5–8] To meaningfully interpret patient narratives,

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (awards K01-AR-079039 and 1F31-AR-084335-01).

[1]Andrew Walker, PhD, Swati Rajwal, MS, Sahithi Lakamana, BSc, Tricia Park, MS, Abeed Sarker, PhD, Selen Bozkurt, PhD, MS: Department of Biomedical Informatics, Emory University School of Medicine, Atlanta, Georgia; [2]Jerik Leung, MPH: Rollins School of Public Health, Emory University, Atlanta, Georgia; [3]Aishwarya Alagappan, MPH, Nathan Le, Titilola Falasinnu, PhD: Division of Immunology and Rheumatology, Stanford University School of Medicine, Stanford, California; [4]Anushka Irani, BM, BCh, PhD: Mayo Clinic, Jacksonville, Florida; [5]Abeed Sarker, PhD: Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25687).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25687.

Address correspondence via email to Selen Bozkurt, PhD, MS, at selen.bozkurt@emory.edu.

Submitted for publication April 8, 2025; accepted in revised form September 22, 2025.

SIGNIFICANCE & INNOVATIONS

- This study demonstrates the utility of applying large language models to extract multidimensional experiences of lupus-related pain in the context of high-volume Reddit online health communities.
- Guided by the biopsychosocial model of pain, we identified common types of pain complaints, management strategies often not discussed in clinical practice, and sociocultural, affective, and nociplastic dimensions of pain.
- Methods used in this study can be applied to explore and monitor emerging themes of patient experiences and behaviors with lupus and other chronic conditions that may not be frequently discussed with providers in clinical settings.

the biopsychosocial model, first applied in the medical domain by Engel,[9] offers a conceptual framework that has been valuable in the complex investigation of chronic pain experience.[10–12] The biopsychosocial model characterizes pain-related suffering as a dynamic, multidimensional experience influenced by intraindividual, interpersonal, and sociocultural factors.[9,10] It also accommodates the evolving role of online communities, in which emotional states and peer interactions can shape the expression and experience of pain.[13]

Recent advances in natural language processing (NLP) and large language models (LLMs) have enabled more sophisticated analyses of psychosocial themes within patient narratives shared online. Bauer et al demonstrated that LLM-based models could detect nuanced indicators of suicidality, including emotional distress, help-seeking behaviors, and social withdrawal, within online health forums.[14] Work by Necaise and Amon in a large chronic pain community on Reddit showed that patients often mirror the emotional tone of their peers, a phenomenon known as sentiment synchrony.[13] These findings suggest that the social-emotional context of peer interactions may influence how patients perceive and express pain, highlighting the importance of understanding pain communication beyond the clinical encounter.

In this study, we applied topic modeling and LLMs to analyze online discussions about pain among individuals with SLE. Guided by the biopsychosocial model of pain,[11] our goal is to identify key themes and patterns that characterize the multidimensional nature of pain as experienced and expressed by patients in real-world settings. By analyzing these discussions, we aim to uncover dimensions of pain that are frequently unvoiced during clinical encounters, including emotional, psychological, and social stressors.[5]

MATERIALS AND METHODS

Data. Reddit posts were collected using the Pushshift.io application programming interface, targeting all publicly available posts from the r/Lupus and r/LupusSupport subreddits from June 9th, 2010, through December 31, 2023.[15] This study was reviewed and classified as exempt by the Emory University Institutional Review Board. All analyses were conducted on aggregated data without user identification or contact, and no verbatim quotations are reported. Data were maintained on secure, access-controlled servers. Consistent with established internet research ethics, measures were implemented to minimize re-identification risk, prevent stigmatizing interpretations, and acknowledge the absence of explicit user consent for research use. The r/Lupus community was created on June 9th, 2010, and r/LupusSupport was created on July 4th, 2020. Reddit does not provide structured demographic information, which limits characterizing the population in our dataset. External surveys suggest that approximately 25% of US adults use Reddit as of 2024, with users younger and more digitally engaged compared with clinic-based cohorts.[16]

Overview of NLP pipeline. Figure 1 summarizes the multistep NLP pipeline used to identify and analyze pain-related content within two Reddit communities (r/Lupus and r/LupusSupport). Each step, including pain-related post identification, topic modeling, LLM summarization, and content analysis, is described in detail subsequently.

Identification of pain-related posts. To identify pain-related posts, we searched for commonly used patient terms describing pain (eg, "aching," "burning," "joint pain") based on a previously published list.[17] Two domain experts (AI and TF) categorized terms into symptom-based groupings relevant to SLE: joint pain, headaches, chest pain, muscle aches, abdominal cramps, and nerve pain, available on GitHub (https://github.com/bozlab/reddit-lupus-pain-nlp). A targeted filtering approach using regular expression matching was applied to identify posts containing these terms. Posts containing at least one term were classified as pain-related and included for analysis. All posts were restricted to English; non-English texts were excluded. We quantified pain-related post frequencies overall and by symptom category to define the analytic dataset.

Topic modeling of pain narratives. To explore naturally occurring themes in patient discussions of pain, we applied BERTopic modeling, an unsupervised machine learning technique, to pain-related posts that had been segmented at the sentence level (n = 15,094 sentences).[18] This approach identifies clusters of co-occurring words and phrases to detect recurring topics in patient narratives. Embeddings were generated with all-MiniLM-L6-v2, reduced using Uniform Manifold Approximation and Projection (n_neighbors = 10, n_components = 2, min_dist = 0.1, cosine metric), and clustered with Hierarchical Density-Based Spatial Clustering of Applications with Noise (min_cluster_size = 50, euclidean metric, excess of mass selection); parameters were chosen to balance coherence and granularity. A CountVectorizer with English stopwords (min_df = 2, ngram_range = 1–4)

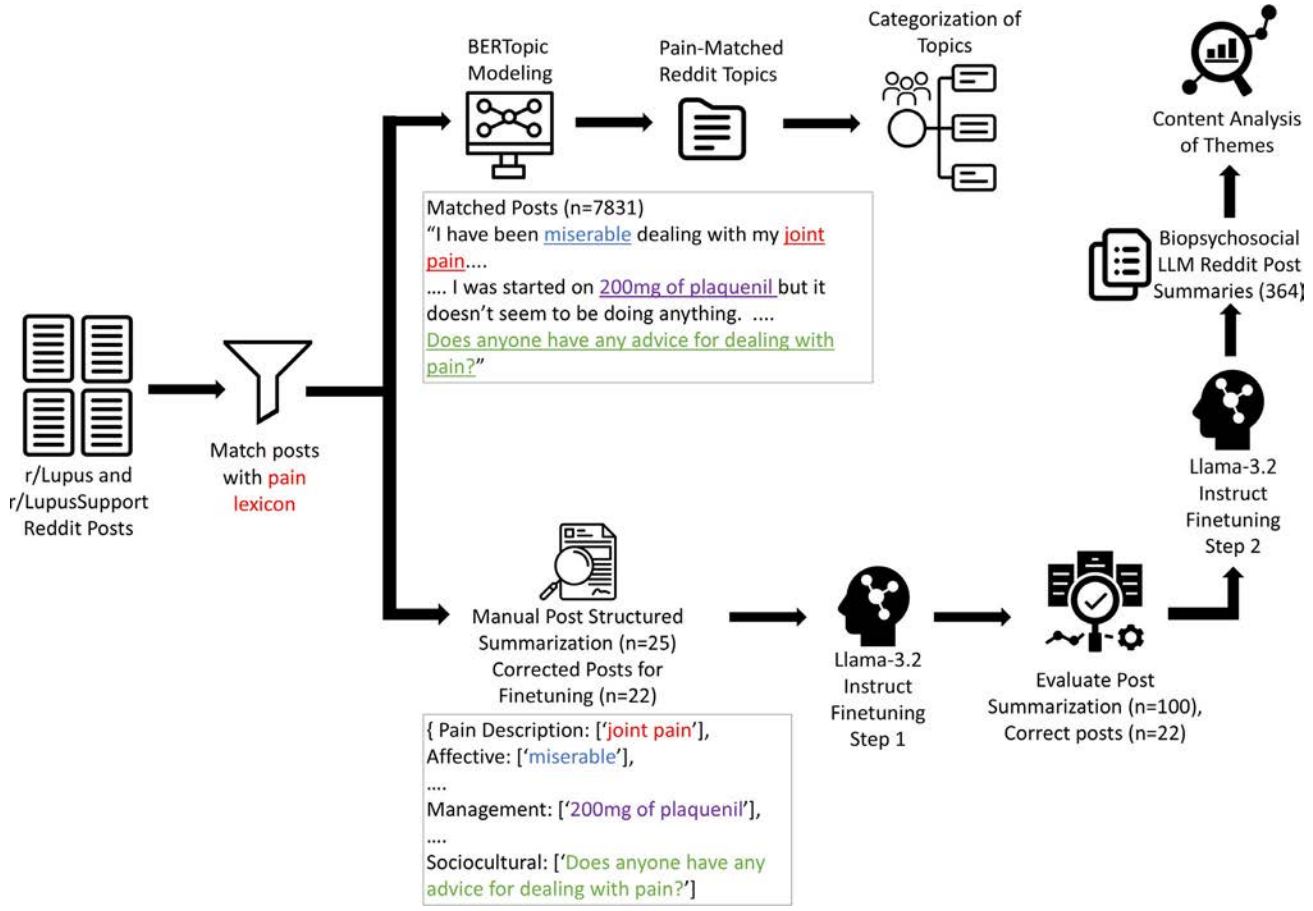


Figure 1. Data analysis pipeline for assessing multidimensional pain experiences in Reddit posts from users in r/Lupus and r/LupusSupport. LLM, large language model. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25687/abstract.

captured frequent and multiword pain phrases. Topic quality was assessed via coherence (70% >0.40) and expert mapping of top words and representative sentences to a predefined biopsychosocial framework (biologic, psychological, social, nociplastic, management) yielding six categories: diagnosis experiences, pain types, pain management, comorbid symptoms, flares, and uncategorized. Full parameter specifications and code for reproducibility are available on GitHub (https://github.com/bozlab/reddit-lupus-pain-nlp).

Manual annotations for biopsychosocial framework. To incorporate a biopsychosocial lens, a random subset of posts (n = 25) was selected from the pool of pain-related posts identified via lexicon matching, restricted to users with self-identified "Diagnosed SLE" user flairs or text and image labels appearing next to usernames in posts and comments in r/Lupus. This subset was manually annotated by three investigators using the biopsychosocial framework to develop the initial codebook and to produce high-quality examples for LLM fine-tuning (Table 1). Content analysis was conducted using structured CSV files rather than qualitative software. Model outputs were exported in JavaScript Object Notation (JSON) format and then organized into spreadsheets for coding. Two investigators independently coded each summary, with discrepancies adjudicated by consensus. This workflow ensured a transparent audit trail, reproducibility of coding decisions, and efficient calculation of frequencies across biopsychosocial dimensions.

LLM fine-tuning and structured summarization. Following manual annotation, the LLaMA 3.2 Instruct 3B original model was instruction fine-tuned using these annotated post-summary pairs, following evidence that small, structured datasets enable effective domain-specific summarization.[19] The model was trained to produce structured summaries in JSON format aligned with the ontology that are readable by humans and parseable for text models.[19] All posts were restricted to English; non-English texts were excluded. We did not apply explicit spell-checking, but the instruction-tuned model demonstrated robustness to common misspellings and colloquial variants, consistent with previous work showing that LLMs trained on web text can normalize noisy input. The model was also prompted to detect negation and contextual absence (eg, "no pain," "pain free") to

Table 1. Pain description dimensions codebook*

Concept	Definition	Examples
Biologic dimension		
Location	Where in the body the pain occurred; can also be general/all over if indicated	"Hands," "joints"
Severity	Author words describing intensity of the pain	"Worst pain I've ever had," "10/10 pain," "mild"
Duration	How long the author describes they have been feeling the pain	"In pain for the last 4 months," "since last week"
Sex/gender	Sex or gender of patient, if mentioned	"Female," "male," "nonbinary"
Age	Author's age, if mentioned	"21," "40," "young," "middle-aged"
Comorbidities	Any comorbidities mentioned in the post alongside lupus	"MCTD," "RA," "fatigue," "loss of vision"
Psychological dimension		
Affective/personal, existential/spiritual	Emotional context of pain, depression, fear, anxiety, stress, or any described emotional sentiment or reactions related to the pain	"I feel miserable now," "so overwhelming," "affecting my anxiety," "driving me crazy," "My life has been taken over by my pain" "I don't feel like myself anymore" "Why is God doing this to me?"
Pain impact on	Identity, threat to self, or personal development, dissatisfaction with life, sense of meaning of life, fear of death, existential concerns, impacts to religious or spiritual beliefs/practices	
Cognitive	Any description of pain impact on mental processes, including coping, catastrophizing, ability to focus, or memory	"Brain fog," "hard to think straight," "difficult to concentrate on anything," "forgetting"
Behavioral	Any changes of a person's day to day actions resulting from pain, including substance use and fear of movement	"I can't eat anymore," "I had to stop working out"
Social dimension		
Functional	Pain impact on ability to complete activities of daily living	"Harder to do the things I want to do," "cannot sit up," "can't eat"
Economic	Pain impact on ability to work, medicolegal, insurance/compensation issues, work environment/job dissatisfaction, absenteeism, presenteeism	"Had to miss work," "can't function at work anymore," "couldn't afford it any longer"
Sociocultural	Followed by any language on how a patient's sociocultural context influenced pain, including stigma, invalidation, or any identified individuals providing support and caregiving, loss of autonomy	"Each time, I visited the doctor, but unfortunately, no action was taken," "How do you deal with the stigma?"
Nociplastic pain	Pain that arises from altered nociception despite 1) no clear evidence of actual or threatened tissue damage that causes peripheral nociceptors activation or 2) no clear evidence of disease or lesion of the somatosensory system causing the pain.	"Fatigue," "brain fog," "pain all over," "nothing seems to be working," "I don't know what to do anymore"
Associated symptoms (often muilti-focal)	Widespread pain, fibromyalgia, vulvodynia, urologic chronic pelvic pain syndrome, chronic tension type headache, chronic low back pain, migraine, temporomandibular disorder	
Pain symptoms	Widespread pain, fibromyalgia, vulvodynia, urologic chronic pelvic pain syndrome, chronic tension type headache, chronic low back pain, migraine, temporomandibular disorder	
Non-pain symptoms	Fatigue, sleep changes, brain fog, uncertainty, overwhelm, helplessness, sensitivity to non-pain stimuli, medication sensitivity and side effects, irritable bowel syndrome	
Pain management	Followed by a list of any methods or medications mentioned by author used to treat the pain as well as any mentions of adherence or barriers/facilitators to access	"Prednisone," "benlysta," "my meds," "hcq," "hydroxychloroquine"

* MCTD, mixed connective tissue disease; RA, rheumatoid arthritis.

avoid false positives (Supplementary Material 1). These behaviors were verified during human evaluation.

Following the initial evaluation of LLM performance, we conducted a second round of model fine-tuning using an expanded set of high-quality training examples (n = 22) that received top scores for structure, accuracy, and coverage and contained no hallucinated content or were corrected after evaluation by consensus coding. This set, combined with the original annotated sample, totaled 47 Reddit posts with structured summaries used for the final round of fine-tuning. The model was fine-tuned again

using the same parameters as in the first round to enhance consistency and completeness across all biopsychosocial dimensions. The updated LLM was then applied to a larger dataset of 317 pain-related posts from r/Lupus, which was restricted to users who self-identified with an "SLE diagnosis" flair. We combined the final summarization sample with the 47 high-quality training samples to ultimately conduct content analysis by AW and JL, iteratively developing codebooks based on consensus coding, on summarizations of 364 posts by users with SLE diagnosis flair. To minimize overfitting, annotated examples were divided into training and evaluation subsets. Fine-tuning was conducted for four epochs with greedy decoding, which emphasized consistency and fidelity to the biopsychosocial ontology rather than maximizing token likelihood. Model performance was monitored on held-out evaluation data, with checkpoints assessed for structured output format, thematic accuracy, and hallucination. Consistent performance across training and held-out sets suggested that the model did not overfit despite the small size of the annotated dataset.

Evaluation of LLM-generated structured summaries. Traditional metrics like precision and recall are insufficient for evaluating the quality of long-form, structured text extracted by LLMs.[20,21] We evaluated outputs using a structured framework for information extraction with LLMs, developed and validated by our NLP team.[22] This framework assesses structure, accuracy, coverage, and hallucination, with expert annotators rating each output on Likert scales to capture semantic adequacy and fidelity.[20,23] Ninety summaries were reviewed by three annotators (AW, JL, AA), with each post scored by two reviewers. Agreement was measured by percent exact and near-exact (±1-point) matches and squared Cohen's kappa values. Full scoring criteria are detailed in Supplementary Material 2.

Content analysis of LLM-summarized pain narratives. The model generated structured summaries for each post, from which representative quotes were extracted and assigned to specific pain dimensions. Two investigators (AW and JL) reviewed these summaries and employed a hybrid coding approach to develop a codebook of thematic categories, drawing on both inductive insights from the data and deductive constructs derived from the biopsychosocial framework. Using this codebook, AW and JL independently applied thematic codes to each quote according to its respective pain dimension. Frequencies of themes were then calculated to identify salient patterns of pain-related suffering expressed within the online community.

RESULTS

The dataset included 31,785 posts from 10,857 unique users, collected between 2010 and 2023 from two lupus-focused subreddits: r/Lupus (42,000 members; top 3% by size) and r/LupusSupport (2,600 members; top 14% by size). Although we do not have demographic data on users available, the top types of flair adopted by Reddit users were "Diagnosed SLE" (n = 694), "Advice" (n = 62), "General" (28), "Medicines" (22), "Caregiver/Loved one" (19), "Diagnosed CLE/DLE" (17), "Seeking Diagnosis" (16), "Venting" (8), and "Physician" (7). Each post

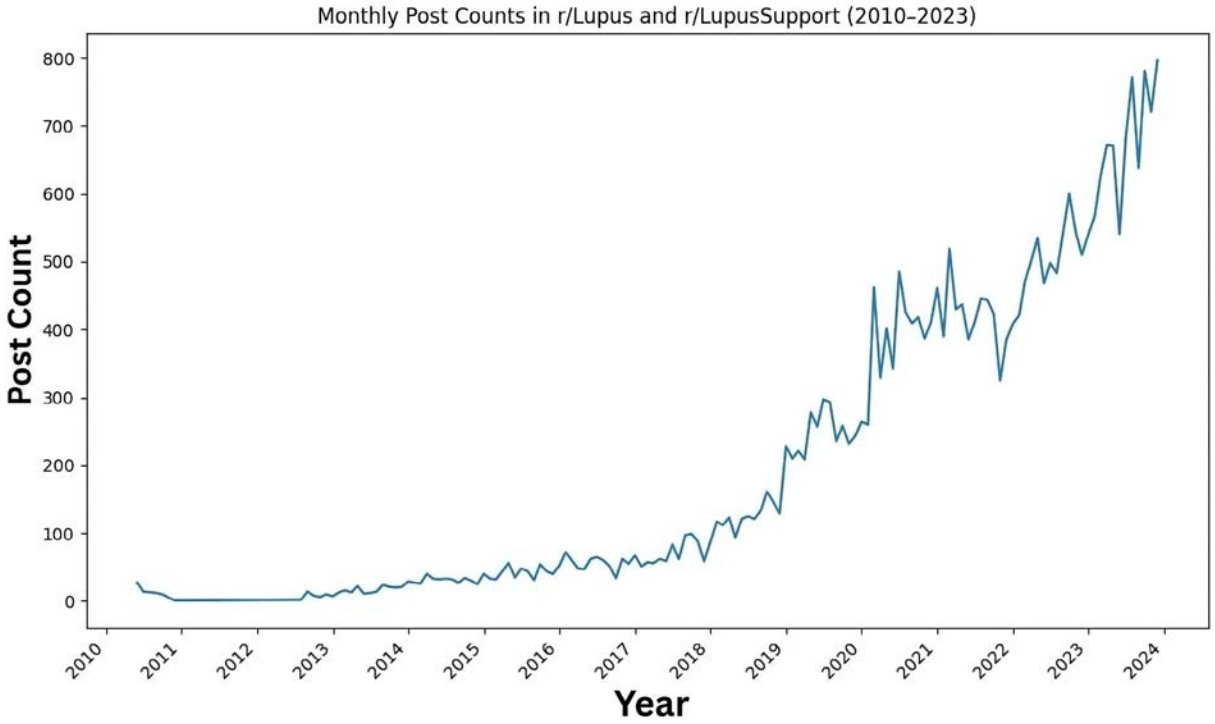


Figure 2. Posts in r/Lupus and r/LupusSupport per month from 2010 to 2023. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25687/abstract.

consisted of an average of 6.9 sentences per post. Figure 2 illustrates the consistent growth in post frequency within r/Lupus.

Identification of pain-related posts. A total of 7,831 posts (24.7%) matched the pain lexicon across r/Lupus and r/LupusSupport. We identified joint pain and arthritis-related terms as the most common lupus symptom mentions (8.0% of all posts), followed by muscle pain/aches (3.8%), headaches (3.3%), chest pain (1.3%) cramps/abdominal pain (1.0%), and spasms/nerve pain (0.7%).

Topic modeling of pain narratives. Topic modeling was applied to 15,094 sentences from 7,831 pain-related posts. Identified themes were grouped into six overarching categories aligned with the biopsychosocial model of pain: diagnosis experiences, pain types, pain management, comorbid symptoms, and flares. A total of 5,009 (33.2%) sentences could not be assigned to a specific topic and were classified as uncategorized. Uncategorized sentences typically involved discussions of joint pains and aches, which may indicate an undercount of joint pain in the overall topic model categorization. Despite the large number of uncategorized sentences (outlier topics), we did not adjust analyses because of the interest in maintaining the coherence in identified topics in line with our biopsychosocial analytic framework, which may have been affected if we reduced the likelihood of outlier classification.[24] Diagnosis experiences of SLE and arthritis were present in 1,728 sentences (11.5%). Figure 3 describes frequencies of categories and topics related to pain types, management, comorbid symptoms, and flares, and Figure 4 describes the top 12 topic frequencies over time.

Within the pain types category (n = 5,668 posts), users described general or widespread pain (n = 1,682, 29.7% of pain experiences category-labeled posts), joint pain (n = 1,385, 24.4%), chest or rib pain (n = 556, 9.8%), and headaches (n = 451, 8.0%). Other pain types included abdominal pain (n = 283, 5.0%), hand and wrist pain (n = 280, 4.9%), kidney pain (n = 168, 3.0%), lower back or spinal pain (n = 127, 2.2%), oral and jaw pain (n = 125, 2.2%), foot or toe pain (n = 108, 1.9%), eye pain (n = 97, 1.7%), scalp or hair-related pain (n = 80, 1.4%), muscle aches (n = 79, 1.4%), menstrual cramps and reproductive pain (n = 78, 1.4%), lymph node pain (n = 61, 1.1%), shoulder pain (n = 54, 1.0%), and hip pain (n = 54, 1.0%). Pain management category posts (n = 687) strategies included over-the-counter medications such as Tylenol and nonsteroidal anti-inflammatory drugs (NSAIDs) (n = 261, 38.0%), glucocorticoids (eg, prednisone, n = 157,22.9%), general pain management discussions (n = 129, 18.8%), disease-modifying therapies like hydroxychloroquine (n = 125, 18.8%), gastrointestinal-related pain interventions (n = 74, 10.8%), and use of cannabis or tetrahydrocannabinol products (n = 70, 10.2%).

Posts referencing comorbid symptoms (n = 1,624 labeled posts) frequently mentioned clusters such as rash, fatigue, brain fog, and fevers (n = 1,247, 76.8% of comorbid labeled posts), as well as sleep disturbances (n = 206, 12.7%), persistent fatigue (n = 86, 5.3%), and muscle spasms (n = 85, 5.2%). Flares were a common context for describing pain (n = 250. 1.7% of all matched sentences). A table including each topic frequency, key words, and relevant matched sentences is available on GitHub (https://github.com/bozlab/reddit-lupus-pain-nlp).

LLM structured summarization results. Following fine-tuning, the LLM model produced 90 of 100 structured summaries that met predefined criteria for accuracy, coverage, and format. Evaluation scores averaged 3.11 of 4 for accuracy and 2.96 of 4 for coverage, with hallucinations flagged in only 8% of

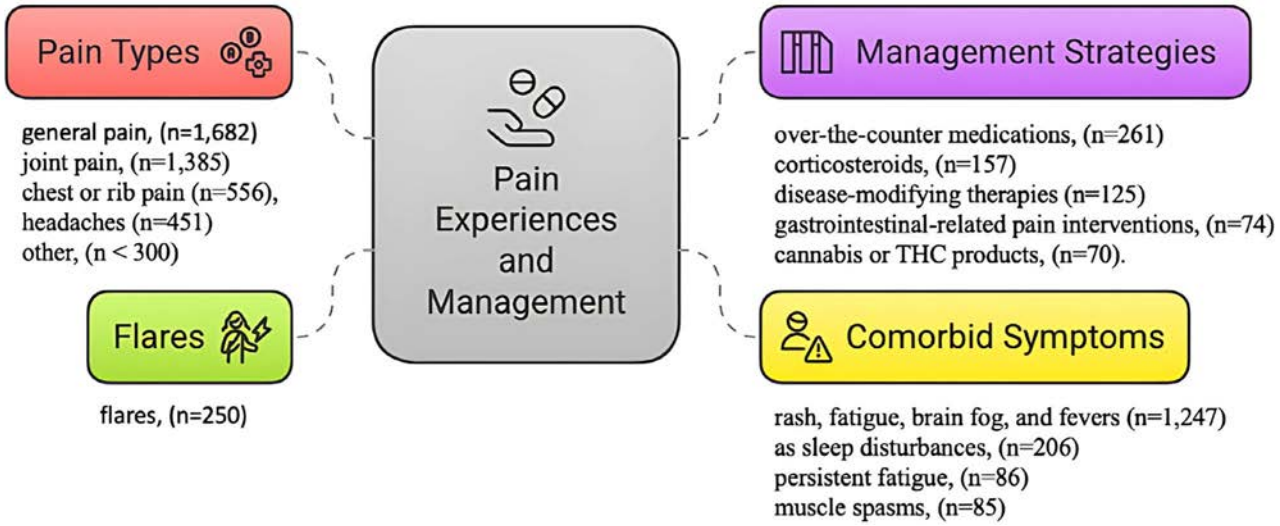


Figure 3. Categories of pain discourse identified from topic modeling in r/Lupus and r/LupusSupport. THC, tetrahydrocannabinol. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25687/abstract.

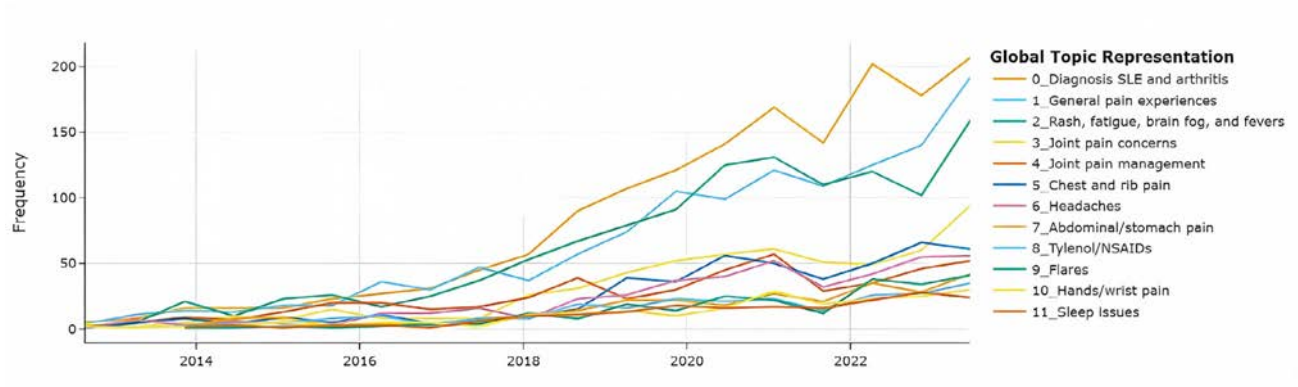


Figure 4. Topic modeling top 12 most prevalent pain-related topics over time on r/Lupus and r/LupusSupport. NSAIDs, nonsteroidal anti-inflammatory drug; SLE, systemic lupus erythematosus.

cases. Exact interrater agreement was moderate, with 92.9% percent interrater agreement on the binary hallucination indicator (Squared-weighted cohen's kappa = .364), 36.4% perfect agreement on accuracy (kappa = .085), and 44.3% perfect agreement on coverage (kappa = .38). Although the use of the squared-weighted kappa attempts to adjust for near-disagreements in ordinal scales,[25] it may not describe the true level of near-agreement, which overall was high, with 87% of accuracy scores and 92% of coverage scores falling within one point between reviewers.

Among the 364 posts summarized, the most frequently reported types of pain were general or widespread pain (n = 305, 83.8%), joint pain (n = 130, 35.7%), and muscle aches (n = 79, 21.7%), with less frequent mentions of headaches (n = 55, 15.1%), chest pain (n = 48, 13.2%), and abdominal pain (n = 34, 9.3%). Structured summaries enabled content analysis of pain-related content extracted from posts across 15 concepts in the biopsychosocial model (Supplementary Material 3). For each concept, we also identified "Incorrect" text extracts, which represented instances in which the model miscategorized a piece of text from the post into the incorrect concept.

Biologic dimension. The biologic dimension was the most frequently annotated across all summaries. Pain location was documented in 314 posts (86.3% of summarized posts), with the most commonly reported areas being "all over" (n = 88, 28.0% of location posts), hands (n = 68, 21.7%), and legs (n = 68, 21.7%). Pain severity appeared in 301 posts (82.7%), often described as severe (n = 146, 48.5% of severity posts), overwhelming (n = 41, 13.6%), or worsening (n = 54, 17.9%). Pain duration was coded in 339 posts (93.1%), with users describing pain lasting for weeks (n = 99, 29.2% of duration posts), months (n = 67, 19.8%), or years (n = 57, 16.8%).

Comorbidities were identified in 263 posts (72.3%), with comorbid symptoms reported in 249 of these posts (94.7%). The most common symptoms included fatigue (n = 57, 21.7%), rashes (n = 36, 13.7%), and nausea (n = 17, 6.5%). Comorbid diseases were mentioned in 100 posts (38.0% of comorbidities posts), with COVID-19 (n = 10, 3.8%), Raynaud phenomenon (n = 6, 2.3%), and arthritis (n = 5, 1.9%) being the most frequently cited.

Demographic dimension. Self-reported gender was identified in 65 posts (17.9%), with the majority identifying as female (n = 62, 95.4% of gender posts). Age was mentioned in 95 posts (26.1%) and categorized using MeSH age groupings. Most users fell within the 19 to 44-year-old adult range (n = 60, 63.2% of age posts).

Psychological dimension. Psychological themes appeared less frequently. Affective content was present in 89 posts (24.5%), with the most common themes including emotional distress (n = 35, 39.9% of affective posts), depression (n = 11, 12.4%), and anxiety (n = 5, 5.6%). Cognitive themes were identified in 24 posts (6.6%), whereas behavioral themes appeared in 11 posts (3.0%). Personal or existential reflections were rare, occurring in only four posts (1.1%).

Sociocultural dimension. Sociocultural content was present in 123 posts (33.8%), primarily describing participation in the subreddit as a source of peer support, advice-seeking, or validation (n = 106, 86.2% of sociocultural posts). Interactions with health care providers were discussed in 35 posts (28.5%). Functional limitations were described in 86 posts (23.6%), with users reporting challenges related to walking, movement, and daily activities. Economic concerns were mentioned in only nine posts (2.5%).

Pain management strategies. Pain management strategies were identified in 280 posts (76.9%), with medications mentioned in 245 of these posts (87.5% of management-related posts). The most frequently cited treatments included prednisone

(n = 92, 32.9%), hydroxychloroquine (n = 91, 32.5%), and NSAIDs (n = 40, 14.3%). Nonpharmacologic strategies were described in 99 posts (35.4%), with common approaches including supplements (n = 24, 8.6%), cannabis/CBD (n = 10, 3.6%), and sun avoidance (n = 12, 4.3%).

Categorization accuracy. On average, 19.83% of annotations were categorized as "Incorrect." The "Sex/Gender" category had the lowest error rate at 0%, whereas the "Severity" category had the highest at 50.8%. Errors in the severity category were often due to repeated pain descriptions without clear indications of severity, ambiguous user phrasing without explicit severity terms, the model inferring severity from adjectives not present in the gold standard, and inconsistent descriptors across multiple severity mentions. These errors could be reduced by incorporating a structured severity lexicon and prompt engineering to enforce explicit severity extraction.

DISCUSSION

Our findings demonstrate that SLE-related pain extends beyond inflammation-driven mechanisms and is shaped by affective, cognitive, social, and nociplastic dimensions. These findings resonate with previous work using data from a population-based registry, which show that disease activity and organ damage account for only part of the variance in pain outcomes,[26] underscoring the influence of psychological, behavioral, and social factors. Building on this evidence, our analysis of patient narratives reveals how nociplastic symptoms (eg, fatigue, brain fog, diffuse pain) and emotional distress often arise independently of conventional disease activity metrics. Notably, users frequently described their pain experiences as disjointed from clinical evaluations or laboratory results, revealing how personal, social, and emotional contexts shape pain perception and impact, emphasizing the need to integrate psychosocial and functional assessments into routine SLE care.[27]

Pain descriptions centered on biologic aspects, with over 86% of patients specifying pain location, most frequently "all over," one of the most commonly cited pain locations in SLE in existing literature.[28,29] Such diffuse presentations are often difficult to reconcile with objective findings on examination or laboratory results. Acknowledging these widespread patterns can reduce diagnostic uncertainty and help rheumatologists differentiate between inflammatory versus nociplastic contributors to pain.

Users also discussed disruptions to daily functioning, including difficulty walking, exercising, or managing basic tasks, emphasizing the central role pain plays in limiting autonomy and participation.[30] In our study, hand-related complaints emerged as a common concern among participants, highlighting the under-researched clinical importance of assessing hand function in SLE.[31,32] Johnsson et al observed that 58% of individuals with SLE demonstrated difficulties on a simple hand function test, compared with only 8% of healthy controls.[31] The clinical presentation of SLE-related joint involvement is notably heterogenous, encompassing a spectrum from noninflammatory arthralgia to more severe forms such as deforming arthropathy (eg, Jaccoud arthropathy) and erosive arthritis.[33] Taken together, these findings underscore the need for routine, nuanced assessment of both widespread and localized joint pain to better capture the functional burden of SLE and guide more tailored, mechanism-informed interventions.

Consistent with prevailing literature, Reddit narratives revealed substantial affective and "silent" burdens of SLE, ranging from depression and anxiety to feelings of hopelessness and invalidation.[34,35] These psychological components are often entangled with nociplastic symptoms such as fatigue, brain fog, sleep disturbances, and widespread pain, which were present in over half of the summarized posts. Such presentations are often challenging to interpret within the conventional markers of disease activity, making diagnosis and treatment more complex.[27] Failing to acknowledge them can result in fragmented care, unnecessary diagnostic testing, and frustrated patients.[36]

Sociocultural themes were prominent, with individuals seeking advice, support, and validation from others. In several posts, users expressed a profound sense of relief and solidarity—captured in statements such as "I'm not alone." These discussions highlight the need for continuous support beyond the clinical setting and suggest that online platforms serve as an important space for emotional coping and peer connection.[13,37] They also emphasize that sociocultural factors, including both health care experiences and broader support systems, meaningfully shape how patients perceive and respond to pain.[38] For clinicians, being attuned to these contextual influences may help explain variations in treatment engagement and foster more empathetic, collaborative relationships.

Our findings on pain management strategies discussed in Reddit closely parallel recent real-world trends in clinical care. A large, national claims-based analysis of over 18,000 patients with SLE observed a steady decline in opioid and NSAID prescriptions since 2014, with increased or stable use of non-opioid modalities such as antidepressants, anticonvulsants, topical agents, and physical therapy.[39] This shift away from opioid-centric management aligns with our observations, in which patients most commonly referenced glucocorticoids (eg, prednisone), hydroxychloroquine, and non-opioid adjunctive agents such as duloxetine, gabapentin, and pregabalin. Reddit users also frequently discussed additional strategies, such as cannabis, supplements, and sun avoidance, many of which are absent from claims data and clinical guidelines.[40] The inclusion of treatments like low-dose naltrexone and alternative therapies may further suggest patients' efforts to address pain through individualized, trial-and-error regimens.[41]

Our findings also demonstrate the value of using advanced NLP tools to structure and scale the analysis of unstructured

patient-reported data. Topic modeling efficiently identified commonly discussed symptom clusters and treatment patterns but lacked the resolution to detect deeper biopsychosocial themes. In contrast, the fine-tuned LLM enabled more granular and concept-aligned extraction of pain narratives across biologic, psychological, social, and nociplastic domains.

This study has several limitations. Diagnostic user flairs on Reddit are self-reported and cannot be clinically verified, introducing the possibility of misclassification bias.[42] This potential misclassification may dilute SLE-specific patterns; however, the biopsychosocial themes identified, particularly those related to nociplastic pain, psychological distress, and sociocultural influences, remain directly relevant to SLE and to rheumatology care more broadly. Previous research indicates that aggregated self-reported data from online health communities can align closely with clinically reported data. The LIFT study, which recruited more than 1,000 individuals with lupus via social media, confirmed lupus diagnoses in 100% of a random subset through medical record review and reported 88% agreement between self-reported and verified medication use.[43] Importantly, r/Lupus is a moderated community with explicit rules that discourage misrepresentation. Community guidelines specify that "Diagnosed" flair should only be used following rheumatologist-confirmed diagnosis, and misuse can result in removal or banning. Users are also instructed not to provide medical advice unless they have a confirmed diagnosis. Although it is not possible to guarantee perfect accuracy, these norms and enforcement mechanisms reduce the likelihood of systematic misrepresentation. Still, compared with clinic-based cohorts, Reddit users tend to be younger, more digitally engaged, and likely to live in North American time zones.[44] Posts also likely disproportionately represented individuals who are motivated to share their experiences online, often those with more severe or poorly controlled symptoms, potentially introducing reporting bias.[45] Future work may also leverage NLP-based demographic inference methods[46] to better contextualize online cohorts relative to clinic-based populations.

Although anonymity on Reddit facilitates candid sharing, it also restricts access to contextual clinical data such as disease duration, flare frequency, treatment regimens, or comorbidity profiles—factors essential for interpreting pain trajectories.[47] Although the fine-tuned LLM performed well on structure and coverage metrics, hallucinations and misclassifications were still observed in a minority of summaries, especially for underrepresented domains such as behavioral, cognitive, existential, and economic factors. These gaps may reflect both model limitations and user hesitancy to disclose sensitive topics online.[48]

Additionally, the LLM training dataset was modest in size (n = 47 posts) and focused on posts with SLE-specific user flairs, which, although high-quality, may limit the model's generalizability to other patient groups or platforms.[49] We mitigated this risk through structured ontology design, manual validation, and schema-constrained outputs. Previous work demonstrates that large pretrained LLMs can adapt effectively to structured information extraction tasks even with limited domain-specific examples.[19] The instruction-tuned LLM we used may also reflect biases present in its pretraining and fine-tuning data, which could influence how pain experiences are categorized or described, particularly for underrepresented cultural, linguistic, or experiential perspectives. To mitigate these risks, we constrained outputs using a biopsychosocial ontology and conducted human evaluation of accuracy, coverage, and hallucination, as recommended by recent bias in NLP mitigation literature.[50]

This study offers several notable strengths that advance the current understanding of pain in SLE from a patient-centered perspective. The use of Reddit as a data source enabled access to candid, unfiltered accounts of patient experiences that may not emerge in structured interviews or electronic health record documentation.[45] Second, the study developed and applied a robust pain ontology grounded in previous literature and expert clinical consensus. This structured framework supported reproducible annotations across biologic, psychological, social, and nociplastic domains—facilitating high-fidelity model training and enabling nuanced content analysis.[9] The inclusion of validation metrics (eg, accuracy, coverage, hallucination rates) for the LLM outputs further strengthens the reliability of findings. Finally, our findings are reinforced by triangulation with existing claims-based studies and registry analyses.[26,39]

In conclusion, this study illustrates the potential of combining unsupervised and LLM-guided NLP approaches to surface-rich, patient-centered perspectives on chronic pain in SLE. These tools can complement clinical assessments by amplifying the patient voice and may be particularly valuable in identifying evolving needs, guiding intervention design, and informing holistic models of care.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Falasinnu T, Nguyen T, Jiang TE, et al. The problem of pain in rheumatology: clinical profiles associated with concomitant diagnoses with chronic overlapping pain conditions. ACR Open Rheumatol. 2022; 4(10):890–896.
2. Pisetsky DS, Eudy AM, Clowse MEB, et al. The categorization of pain in systemic lupus erythematosus. Rheum Dis Clin North Am. 2021; 47(2):215–228.

3. Goudman L, De Smedt A, Moens M. Social media and chronic pain: what do patients discuss? J Pers Med. 2022;12(5):797.
4. Bell L, Fordham B, Mumtaz S, et al. Using natural language processing and social media data to understand the lived experience of people with fibromyalgia. Healthcare (Basel). 2024;12(24):2511.
5. Harter C, Ness M, Goldin A, et al. Exploring chronic pain and pain management perspectives: qualitative pilot analysis of web-based health community posts. JMIR Infodemiology. 2023;3:e41672.
6. US Food and Drug Administration. Patient-focused drug development: collecting comprehensive and representative input. Published January 26, 2021. Accessed March 27, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-collecting-comprehensive-and-representative-input
7. Dreyfus B, Chaudhary A, Bhardwaj P, et al. Application of natural language processing techniques to identify off-label drug usage from various online health communities. J Am Med Inform Assoc. 2021;28(10):2147–2154.
8. Best Practices for Food and Drug Administration Staff in the Postmarketing Safety Surveillance of Human Drug and Biological Products. Final Document; Availability. *Federal Register*. Published January 26, 2024. Accessed March 27, 2025. https://www.federalregister.gov/documents/2024/01/26/2024-01584/best-practices-for-food-and-drug-administration-staff-in-the-postmarketing-safety-surveillance-of
9. Engel GL. The need for a new medical model: a challenge for biomedicine. Science. 1977;196(4286):129–136.
10. Gatchel RJ, Peng YB, Peters ML, et al. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull. 2007;133(4):581–624.
11. Miaskowski C, Blyth F, Nicosia F, et al. A biopsychosocial model of chronic pain for older adults. Pain Med. 2020;21(9):1793–1805.
12. Black NM, Sullivan SJ, Mani R. A biopsychosocial understanding of lower back pain: content analysis of online information. Eur J Pain. 2018;22(4):728–744.
13. Necaise A, Amon MJ. Peer support for chronic pain in online health communities: quantitative study on the dynamics of social interactions in a chronic pain forum. J Med Internet Res. 2024;26(1):e45858.
14. Bauer B, Norel R, Leow A, et al. Using large language models to understand suicidality in a social media-based taxonomy of mental health disorders: linguistic analysis of Reddit posts. JMIR Ment Health. 2024;11:e57234.
15. NCRI. Accessed March 24, 2025. https://pushshift.io/signup
16. Social MFS. Pew Research Center. Published November 13, 2024. Accessed August 30, 2025. https://www.pewresearch.org/internet/fact-sheet/social-media/
17. Chaturvedi J, Mascio A, Velupillai SU, et al. Development of a lexicon for pain. Front Digit Health. 2021;3:778305.
18. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv. Preprint posted online March 11, 2022. https://doi.org/10.48550/arXiv.2203.05794
19. Dagdelen J, Dunn A, Lee S, et al. Structured information extraction from scientific text with large language models. Nat Commun. 2024;15(1):1418.
20. Subbiah M, Zhang S, Chilton LB, et al. Reading subtext: evaluating large language models on short story summarization with writers. Trans Assoc Comput Linguist. 2024;12:1290–1310.
21. Castilho S. On the same page? Comparing inter-annotator agreement in sentence and document level human machine translation evaluation. In: Barrault L, Bojar O, Bougares F, et al., eds. Proceedings of the Fifth Conference on Machine Translation. Association for Computational Linguistics; 2020:1150–1159. Accessed March 28, 2025.
22. Das S, Ge Y, Guo Y, et al. Two-layer retrieval-augmented generation framework for low-resource medical question answering using reddit data: proof-of-concept study. J Med Internet Res. 2025;27(1):e66220.
23. Das S, Ge Y, Guo Y, et al. Two-layer retrieval augmented generation framework for low-resource medical question-answering: proof of concept using Reddit data. arXiv. 2024. Accessed May 31, 2024. http://arxiv.org/abs/2405.19519
24. Grootendorst M. Outlier reduction - BERTopic. 2024. Accessed August 30, 2025. https://maartengr.github.io/BERTopic/getting_started/outlier_reduction/outlier_reduction.html
25. Li M, Gao Q, Yu T. Kappa statistic considerations in evaluating inter-rater reliability between two raters: which, when and context matters. BMC Cancer. 2023;23(1):799.
26. Falasinnu T, Drenkard C, Bao G, et al. The problem of pain in systemic lupus erythematosus: an explication of the role of biopsychosocial mechanisms. J Rheumatol. 2021;48(8):1279–1289.
27. Eudy AM, Clowse ME, Corneli A, et al. The Type 1 & 2 systemic lupus erythematosus model: perspectives of people living with systemic lupus erythematosus. Lupus. 2024;33(3):266–272.
28. Iannuccelli C, Spinelli FR, Guzzo MP, et al. Fatigue and widespread pain in systemic lupus erythematosus and Sjögren's syndrome: symptoms of the inflammatory disease or associated fibromyalgia? Clin Exp Rheumatol. 2012;30(6)(S74):117–121.
29. Jiang TE, Mackey S, Darnall BD, et al. The problem of pain in systemic lupus erythematosus: a comprehensive analysis of pain distribution using the CHOIR body map and PROMIS measures. Lupus. 2025;34(1):47–56.
30. Brennan F, Carr DB, Cousins M. Pain management: a fundamental human right. Anesth Analg. 2007;105(1):205–221.
31. Malcus Johnsson P, Sandqvist G, et al. Hand function and performance of daily activities in systemic lupus erythematosus: a clinical study. Lupus. 2015;24(8):827–834.
32. Malcus Johnsson P, Sandqvist G, Bengtsson A, et al. Hand function and performance of daily activities in systemic lupus erythematosus. Arthritis Rheum. 2008;59(10):1432–1438.
33. Marone A, Tang W, Kim Y, et al. Evaluation of SLE arthritis using frequency domain optical imaging. Lupus Sci Med. 2021;8(1):e000495.
34. Fonseca R, Bernardes M, Terroso G, et al. Silent burdens in disease: fatigue and depression in SLE. Autoimmune Dis. 2014;2014:790724.
35. Elefante E, Tani C, Stagnaro C, et al. Impact of fatigue on health-related quality of life and illness perception in a monocentric cohort of patients with systemic lupus erythematosus. RMD Open. 2020;6(1):e001133.
36. Eudy AM, Clowse MEB, Corneli A, et al. Perspectives of rheumatologists on the type 1 and 2 systemic lupus erythematosus model. ACR Open Rheumatol. 2024;6(12):865–870.
37. Goren I, Sharvit G, Godny L, et al. Exploring popular social media networks for patients with inflammatory bowel diseases: insights for better care. J Clin Gastroenterol. 2022;56(3):e203–e208.
38. Rogger R, Bello C, Romero CS, et al. Cultural framing and the impact on acute pain and pain services. Curr Pain Headache Rep. 2023;27(9):429–436.
39. Falasinnu T, Lu D, Baker MC. Annual trends in pain management modalities in patients with newly diagnosed autoimmune rheumatic diseases in the USA from 2007 to 2021: an administrative claims-based study. Lancet Rheumatol. 2024;6(8):e518–e527.
40. Guillouard M, Authier N, Pereira B, et al. Cannabis use assessment and its impact on pain in rheumatologic diseases: a systematic review and meta-analysis. Rheumatology (Oxford). 2021;60(2):549–556.
41. Aitcheson N, Lin Z, Tynan K. Low-dose naltrexone in the treatment of fibromyalgia: a systematic review and narrative synthesis. Aust J Gen Pract. 2023;52(4):189–195.

42. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med. 2020;3(1):43.
43. Cheeks A, Borisov L, Warren K, et al. FRI0606 the LIFT study's direct-to-patient design provides reliable self-reported data from lupus patients. Ann Rheum Dis. 2018;77(S2):826–827.
44. Choudhury MD, Gamon M, Counts S, et al. Predicting depression via social media. Proc Int AAAI Conf Web Soc Media. 2013;7(1): 128–137.
45. Guntuku SC, Yaden DB, Kern ML, et al. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci. 2017;18:43–49.
46. Yang YC, Al-Garadi MA, Love JS, et al. Can accurate demographic information about people who use prescription medications nonmedically be derived from Twitter? Proc Natl Acad Sci USA. 2023;120(8): e2207391120.
47. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol. 2016; 9:77–82.
48. Morstatter F, Pfeffer J, Liu H, et al. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. Proc Int AAAI Conf Web Soc Media. 2013;7(1):400–408.
49. Devlin J, Chang MW, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Burstein J, Doran C, Solorio T, eds. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *Volume 1: (Long and Short Papers)*. Association for Computational Linguistics; 2019:4171–4186.
50. Raza S, Shaban-Nejad A, Dolatabadi E, et al. Exploring bias and prediction metrics to characterise the fairness of machine learning for equity-centered public health decision-making: a narrative review. IEEE Access. 2024;12:180815–180829.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 134–140
DOI 10.1002/acr.25571

BRIEF REPORT

Prevalence of Pain and Factors Associated With Pain in Patients With Idiopathic Inflammatory Myopathies

Didem Saygin,[1] Anne-Marie Malfait,[2] Ethan M. Ritz,[1] Kristin Wipfler,[3] Kaleb Michaud,[4] and Yvonne C. Lee[5]

Objective. Pain is an often overlooked and understudied symptom in idiopathic inflammatory myopathies (IIM). In this study, our goal is to examine the prevalence of pain and factors associated with pain in adults with IIM.

Methods. FORWARD is a US-based prospective registry of adults with rheumatic diseases recruited from rheumatology clinics. Participants had a physician diagnosis of IIM and completed questionnaires on pain, fatigue, physical function, disease activity, and sociodemographic variables. Pain prevalence was examined in demographic and disease subgroups. Regression models were performed to identify any factors associated with higher levels of pain (>3 on visual analog scale). The relationship between pain and outcome variables was examined through Pearson correlation, the *t*-test, and the chi square test.

Results. A total of 189 patients with IIM (mean age ± SD 55 ± 14 years, 78% women) from the FORWARD databank were included in the study. Approximately 86% reported pain with an average of 3.5 on a 10-cm visual analog scale. One in four patients were taking opioids. Approximately 62% to 63% reported joint and muscle pain, respectively. Pain prevalence was similar across subgroups of disease subtype, sex, race, smoking status, and obesity. Patient global disease activity was significantly associated with higher pain levels in multivariable models. Pain was significantly associated with worse physical function, fatigue, quality of life, and health satisfaction.

Conclusion. Pain is prevalent and associated with poor outcomes in patients with IIM. Pain is closely linked with patient global disease activity. Results highlight the critical need to better understand the pain experienced by these patients to best address their needs.

INTRODUCTION

Idiopathic inflammatory myopathies (IIM) are a heterogenous group of autoimmune diseases characterized by skeletal muscle inflammation. The cardinal symptom of IIM is skeletal muscle weakness leading to functional limitations in the majority of the patients. Therefore, outcomes used to assess patients with IIM in clinical practice and therapeutic trials naturally focus on muscle weakness and physical function domains. With a recent shift toward better understanding the lived experiences of patients with disease, other disease symptoms that are important to patients are being increasingly recognized. Because IIM was traditionally believed to cause a painless weakness, pain has long been overlooked as a symptom. However, approximately 70% to 80% of patients with IIM report pain that interferes with their daily activities.[1,2] In addition, several focus groups and survey studies revealed that pain is one of the most important symptoms for patients with IIM.[3] Pain is also one of the leading contributors to discordance between patient- and physician-reported myositis disease activity because patients are more likely to consider pain as a manifestation of their disease than providers are.[4] This observation suggests that pain could be frequently dismissed or possibly misattributed to other causes in evaluation of myositis disease activity by health care providers.

Dr Saygin's work was supported by the Rheumatology Research Foundation Scientist Development Award, the Rukel Family Foundation, the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), NIH (C-COMP Pilot Grant P30-AR-079206), and the RUSH to Progress Pilot Award. Dr Malfait's work was supported by the NIAMS, NIH (grants R01-AR-060364, R01-AR-064251, and P30-AR-079206). Dr Michaud's work was supported by the Rheumatology Research Foundation Innovative Research Award. Dr Lee's work was supported by the NIH (grants K24-AR-080840, R01-AR-064850, and P30-AR-072579) and a Rheumatology Research Foundation Innovative Research Award.

[1]Didem Saygin, MD, Ethan M. Ritz, MS: Rush University Medical Center, Chicago, Illinois; [2]Anne-Marie Malfait, MD, PhD: Rush University Medical Center and Chicago Center on Musculoskeletal Pain, Chicago, Illinois; [3]Kristin Wipfler, PhD: Chicago Center on Musculoskeletal Pain, Chicago, Illinois; [4]Kaleb Michaud, PhD: FORWARD, The National Databank for Rheumatic Diseases, Wichita, Kansas, and University of Nebraska Medical Center, Omaha; [5]Yvonne C. Lee, MD, MMSc: Northwestern University, Chicago, Illinois.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25571.

Address correspondence via email to Didem Saygin, MD, at didem_saygin@rush.edu.

Submitted for publication February 6, 2025; accepted in revised form May 1, 2025.

SIGNIFICANCE & INNOVATIONS

- Contrary to the common belief that myositis is a painless disease, a majority of patients with myositis report muscle and joint pains.
- One in four patients with inflammatory myopathies uses opioids for their pain.
- Patients with myositis who have pain have significantly worse physical function, fatigue, quality of life, and health satisfaction.
- Patients with myositis may perceive a close link between their pain and their myositis disease activity.

Rigorous research is urgently needed to better understand the pain experienced by patients with IIM, which can, in turn, improve the care provided to these patients. However, a systematic literature review highlighted that despite the high prevalence of pain, studies focusing on pain in IIM are scarce.[1] In this study, we aim to (1) assess the prevalence and intensity of pain experienced by patients with IIM overall, as well as in demographic and disease subgroups; (2) identify sociodemographic and disease-related factors associated with pain; and (3) explore the relationship between pain and important health-related outcomes in patients with IIM.

PATIENTS AND METHODS

Study population. This was a cross-sectional observational study using the FORWARD databank. FORWARD is a long-term prospective registry of adults with rheumatic diseases in the United States.[5] Participants in FORWARD are primarily recruited from rheumatology clinics in the United States and complete comprehensive semiannual questionnaires regarding their health status, including physical function, pain, fatigue, health-related quality of life, and visits with various health care providers. These questionnaires include demographic variables; body mass index (BMI); education level; annual income; history of smoking; disease symptom duration (years since symptom onset of IIM); disease subtype (dermatomyositis [DM], polymyositis or immune-mediated necrotizing myopathy [PM/IMNM], overlap myositis, inclusion body myositis [IBM], and unspecified); 10-cm visual analog scales (VAS) for pain and fatigue; health satisfaction (Likert scale, 0–4); patient global disease activity (0–10); the Health Assessment Questionnaire II (HAQ-II, 0–3); the Short Form 36 (SF-36, 0–100); the Rheumatic Disease Comorbidity Index (0–9); self-reported current depression and anxiety; history of osteoarthritis, Raynaud phenomenon, hypertension, myocardial infarction, diabetes mellitus, cancer, renal disorder, pulmonary disorder, fracture, gastrointestinal (GI) disorder, and cardiac condition; and current medications (nonsteroidal anti-inflammatory drugs [NSAIDs], opioids, glucocorticoids, and immunosuppressive or modulatory therapy). Participants included in this study were those who had a physician-confirmed diagnosis of IIM. The FORWARD registry procedures were approved by the Via Christi Institutional Review Board, and all participants provided consent to participate.

Statistical analyses. Participant characteristics at cohort entry were reported using descriptive statistics. For the pain VAS, participants were asked the following: "How much pain have you had because of your illness in the past week?". Participants were asked to mark the scale based on what best described their pain from 0 (no pain) to 10 (severe pain). Pain prevalence was defined as the proportion of patients who had >0 on a 10-cm VAS for pain. Pain prevalence was compared using the chi square test across groups (DM vs non-DM, women vs men, White individuals vs those of other races, patients with a history of smoking vs never smokers, and those with obesity vs those without obesity [BMI ≥30 vs <30]).

The variables were divided into factors that could contribute to pain (age, sex, race, BMI, education level, annual income, history of smoking, disease symptom duration, disease subtype, comorbidities, and the Rheumatic Disease Comorbidity Index) and outcomes that could be associated with pain (fatigue, patient global disease activity, physical function [HAQ-II], health-related quality of life [SF-36 physical and mental component scores], and health satisfaction).

To examine the relationship between factors hypothesized to contribute to pain and the presence of significant pain (categorical variable, ≤3 vs >3 on 10-cm VAS), logistic regression models were constructed; odds ratios (for one-unit increase), confidence intervals, and P values were reported. The 3-cm cutoff was selected based on literature showing that pain levels >3 are associated with interference with daily activities and at least a mild pain on categorical assessment.[6] Univariate models were constructed, and variables significantly associated with pain ($P < 0.05$) were considered for inclusion in multivariable models. Individual comorbidities were considered for inclusion as opposed to the Rheumatic Disease Comorbidity Index because we thought that identifying specific comorbidities would be more likely to yield actionable results. Multicollinearity between predictor variables was assessed using a correlation matrix with Pearson correlation for continuous variables and a contingency table with the chi square test for categorical variables. $P < 0.05$ was considered as statistically significant. To address the significant correlation between depression and anxiety ($P < 0.0001$), only depression was included in the models. The maximum number of independent variables was capped at 10 to avoid overfitting. A sensitivity analysis was performed, excluding patient global assessment as an independent variable in the multivariable regression model.

To explore the relationship between pain and other important health-related outcomes, we used (1) Pearson correlations for

continuous variables (10-cm pain intensity VAS and other health-related outcomes) and (2) the *t*-test and chi square test for categorical outcome variables in patients with pain scores of ≤3 versus >3 on the 10-cm VAS. Correlations were interpreted as weak for r between 0 and 0.3, moderate for r between 0.3 and 0.7, and strong for r >0.7. Analyses were conducted using a complete case approach, whereby only participants with no missing data on variables of interest were included. Statistical analyses were conducted using GraphPad Prism 10 and R (version 4.4.1).

RESULTS

Study population. A total of 210 patients had an IIM diagnosis in the FORWARD databank. After the patients with missing data on pain were excluded (n = 21), 189 patients (average age ± SD 54.7 ± 13.7 years, 78.3% female) were included in the study (Table 1). The patients with missing data (<10% of the total cohort) were slightly older (average age of 60.5 years) and included fewer women (61.9%). Racial groups included White, non-Hispanic (n = 145), Black (n = 14), Hispanic (n = 8), Asian or Pacific Islander (n = 3), American Indian or Alaska Native (n = 1), and unknown (n = 18). A total of 68 patients had DM, 65 had PM/IMNM, 48 had unspecified inflammatory myopathy, 6 had IBM, and 2 had overlap myositis with an average symptom duration of 8.6 years (SD ±7.4). Approximately 53.4% of the patients were taking glucocorticoids, and 71.4% were taking an immunosuppressive or modulatory agent.

Pain prevalence and intensity in patients with IIM. Of 189 patients with IIM, 85.6% (n = 162) reported any pain (>0 on VAS) and 44.9% (n = 85) reported pain >3 on the

Table 1. Characteristics of the study population (patients with pain levels ≤3 and >3 on VAS)*

	Pain ≤3 (n = 104)	Pain >3 (n = 85)	All (N = 189)
Sociodemographic variables			
Age, mean (SD), y	57.5 (14.3)	51.2 (12.2)	54.7 (13.7)
Sex (female), n (%)	80 (76.9)	68 (80.0)	148 (78.3)
Race (White), n (%)	88 (84.6)	57 (67.1)	145 (76.7)
Smoking (ever smoker), n (%)	38 (36.5)	48 (56.5)	86 (45.5)
BMI, mean (SD)	27.3 (6.0)	29.9 (7.3)	28.4 (6.7)
Education level, mean (SD), y	14.5 (2.1)	13.6 (2.2)	14.1 (2.2)
Annual income, mean (SD), ×$1,000	60.9 (30.9)	45.8 (28.9)	53.9 (30.8)
Symptom duration, mean (SD), y	9.0 (8.1)	7.9 (6.6)	8.6 (7.4)
Disease subtype (DM), n (%)	40 (38.4)	28 (32.9)	68 (36.5)
Outcomes, mean (SD)			
Fatigue (0–10)	3.0 (2.7)	6.9 (2.6)	4.7 (3.3)
Patient global (0–10)	2.6 (2.3)	5.6 (2.2)	3.9 (2.7)
HAQ-II (0–3)	0.7 (0.7)	1.4 (0.5)	1.1 (0.7)
Patient Activity Scale (0–10)	2.1 (1.5)	5.6 (1.6)	3.7 (2.3)
SF-36 PCS (0–100)	41.1 (10.4)	30.0 (8.5)	36.5 (11.1)
SF-36 MCS (0–100)	50.5 (10.2)	42.4 (12.5)	47.1 (11.9)
Health satisfaction (0–4)	2.7 (1.2)	1.4 (1.2)	2.1 (1.3)
Comorbidities			
Rheumatic Disease Comorbidity Index (0–9), mean (SD)	1.9 (1.5)	2.4 (1.8)	2.1 (1.7)
Fibromyalgia, n (%)	12 (11.5)	44 (51.8)	56 (29.6)
Osteoarthritis, n (%)	14 (13.5)	15 (17.6)	29 (15.3)
Raynaud phenomenon, n (%)	20 (19.2)	15 (17.6)	35 (18.5)
Depression, n (%)	16 (15.4)	33 (38.8)	49 (25.9)
Anxiety, n (%)	12 (11.5)	26 (30.6)	38 (20.1)
Hypertension, n (%)	52 (50.0)	42 (49.4)	94 (49.7)
Myocardial infarction, n (%)	3 (2.9)	7 (8.2)	10 (5.3)
Diabetes mellitus, n (%)	8 (7.7)	17 (20.0)	25 (13.2)
History of cancer, n (%)	18 (17.3)	11 (12.9)	29 (15.3)
History of renal disorder, n (%)	8 (7.7)	11 (12.9)	19 (10.1)
History of pulmonary disorder, n (%)	25 (24.0)	26 (30.6)	51 (26.9)
History of fracture, n (%)	9 (8.7)	13 (15.3)	22 (11.6)
History of GI disorder, n (%)	44 (42.3)	49 (57.6)	93 (49.2)
Cardiac condition, n (%)	19 (18.3)	17 (20.0)	36 (19.0)
Medications, n (%)			
NSAID	23 (22.1)	32 (37.6)	55 (29.1)
Opioid	13 (12.5)	30 (35.3)	43 (22.8)
Glucocorticoid	63 (60.6)	38 (44.7)	101 (53.4)
Immunosuppressive/modulatory therapy	72 (69.2)	63 (74.1)	135 (71.4)

* BMI, body mass index; DM, dermatomyositis; GI, gastrointestinal; HAQ-II, Health Assessment Questionnaire II; MCS, mental component summary; NSAID, nonsteroidal anti-inflammatory drug; PCS, physical component summary; SF-36, Short Form 36; VAS, visual analog scale.

VAS (Table 1). The average pain level was 3.5 (SD ±2.9) of 10 in patients with IIM. Approximately, 63.2% reported joint pain, and 61.7% reported muscle pain. A total of 29.6% of patients reported a diagnosis of fibromyalgia. Approximately 29.1% and 22.8% of patients were taking NSAIDs and opioids, respectively. The pain prevalence was comparable in patients with DM versus non-DM (87.5% vs 82.4%), women versus men (85.8% vs 86.5%), White individuals versus those of other races (84.8% vs 84.6%), those with a history of smoking versus never smokers (83.7% vs 87.4%), those with obesity versus those without obesity (86.2% vs 86.0%), and early versus late disease (88.0% vs 84.3%).

Factors associated with pain levels >3 in patients with IIM. Of all the variables tested, the following variables were associated with increased odds of having pain levels >3 in the univariate models: increased patient global disease activity, younger age, races other than White, smoking history, higher BMI, lower education level, higher Rheumatic Disease Comorbidity Index score, and history of depression, anxiety, diabetes mellitus, and GI disorder (Table 2). The multivariable model included the following nine variables: patient global disease activity, age, race, smoking history, BMI, education level, depression, diabetes mellitus, and GI disorder. When all the nine variables were included in the multivariable model, patient global disease activity was the only variable significantly associated with increased odds of having pain levels >3. When the model was repeated without including patient global disease activity, smoking history was the only variable that was significantly associated with having higher odds of pain levels >3.

Relationship between pain and outcome variables. Pain intensity had moderate positive correlations with fatigue and moderate negative correlations with physical function (HAQ-II), health-related quality of life (SF-36 physical and mental component scores) and health satisfaction (Figure 1). Patients who reported pain levels >3 had significantly worse mean ± SD fatigue (6.9 ± 2.6 vs 3.0 ± 2.7), HAQ-II (1.4 ± 0.5 vs 0.7 ± 0.7), and SF-36 physical (30 ± 8.5 vs 41.1 ± 10.4) and mental component scores (42.4 ± 12.5 vs 50.5 ± 10.2) and health satisfaction (1.4 ± 1.2 vs 2.7 ± 1.2) than those with pain levels ≤3 (Table 1).

DISCUSSION

In this observational cohort study, the majority of patients with IIM reported pain, with approximately 45% reporting pain levels greater than 3 out of 10. One in four patients with IIM were taking opioids. Higher patient global disease activity was

Table 2. Univariate and multivariable logistic regression models (with and without patient global disease activity) to identify the factors associated with pain levels >3 in patients with idiopathic inflammatory myopathies*

	Univariate logistic regression model			Multivariable logistic regression model with patient global disease activity			Multivariable logistic regression model without patient global disease activity		
	OR	CI	*P* value	OR	CI	*P* value	OR	CI	*P* value
Patient global disease activity	**1.72**	**1.48–2.03**	**<0.0001**	**1.51**	**1.28–1.81**	**<0.0001**	–	–	–
Sociodemographic variables									
Age, y	**0.97**	**0.94–0.99**	**0.002**	0.97	0.94–1.00	0.07	0.97	0.95–1.00	0.06
Symptom duration, y	0.98	0.94–1.02	0.3	–	–	–	–	–	–
Sex (female)	1.12	0.54–2.33	0.8	–	–	–	–	–	–
Race (White)	**0.40**	**0.17–0.94**	**0.04**	0.68	0.21–2.19	0.5	0.67	0.23–1.91	0.4
Smoking (yes)	**2.25**	**1.26–4.07**	**0.007**	1.69	0.72–4.00	0.2	2.20	**1.03–4.78**	**0.04**
BMI	**1.06**	**1.01–1.12**	**0.01**	1.03	0.96–1.09	0.4	1.05	0.99–1.11	0.09
Education level, y	**0.81**	**0.69–0.94**	**0.007**	0.87	0.70–1.09	0.2	0.82	0.67–1.00	0.055
Annual income (×$1,000)	1.00	0.99–1.02	0.09	–	–	–	–	–	–
Non-DM (vs DM)	0.80	0.44–1.46	0.4	–	–	–	–	–	–
Comorbidities									
Rheumatic Disease Comorbidity Index	**1.20**	**1.01–1.44**	**0.04**	–	–	–	–	–	–
Osteoarthritis	1.38	0.62–3.07	0.4	–	–	–	–	–	–
Depression	**4.13**	**2.07–8.52**	**<0.0001**	1.43	0.52–3.94	0.5	2.14	0.88–5.30	0.09
Anxiety	**3.93**	**1.85–8.73**	**0.0005**	–	–	–	–	–	–
Hypertension	0.98	0.55–1.73	0.9	–	–	–	–	–	–
Myocardial infarction	3.02	0.82–14.36	0.1	–	–	–	–	–	–
Diabetes mellitus	**3.00**	**1.26–7.72**	**0.02**	2.45	0.66–9.53	0.2	2.44	0.78–8.06	0.13
Cancer	0.71	0.31–1.58	0.4	–	–	–	–	–	–
Renal disorder	1.78	0.69–4.82	0.2	–	–	–	–	–	–
Pulmonary disorder	1.39	0.73–2.66	0.3	–	–	–	–	–	–
Gastrointestinal disorder	**1.86**	**1.04–3.33**	**0.04**	1.62	0.69–3.82	0.3	1.60	0.75–3.42	0.23
Cardiac condition	1.11	0.54–2.32	0.7	–	–	–	–	–	–

* Bold values indicate statistically significant results with $P < 0.005$. BMI, body mass index; CI, confidence interval; DM, dermatomyositis; OR, odds ratio.

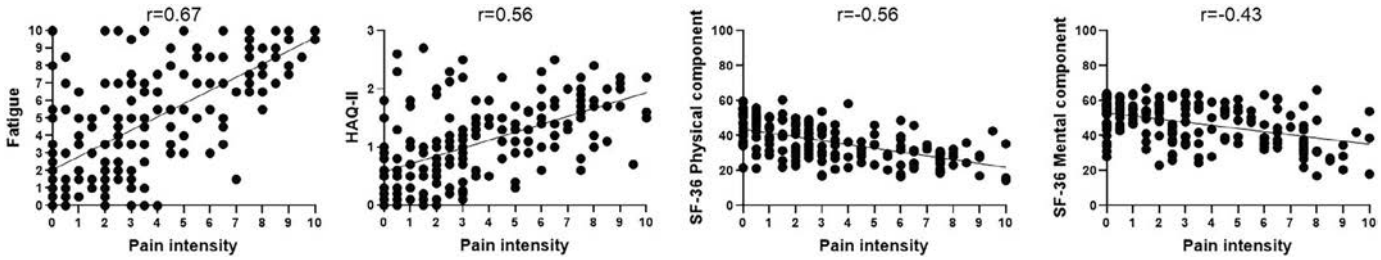


Figure 1. Correlations between pain intensity and outcome variables in patients with idiopathic inflammatory myopathies. HAQ-II, Health Assessment Questionnaire; SF-36, Short Form 36.

significantly linked to having pain levels >3. Higher pain levels were associated with functional disability and worse fatigue, health-related quality of life, and health satisfaction. Overall, these findings demonstrate the high prevalence and significant burden of pain among patients with IIM and emphasize the importance for further studies to understand myositis-related pain. High rates of opioid use highlight the need for alternative pain management approaches in these patients.

The average pain level in patients with IIM was 3.5 on the 10-cm VAS. Although the validity of this measure in IIM has not been established, other studies using the 10-cm VAS or numeric rating scale have reported average pain levels between 2 and 3 among patients with IIM.[1] Comparable to our results, the frequency of patients with pain levels >3 of 10 was 45% in another US-based study.[7] Sex was not significantly associated with pain levels, and pain prevalence was similar between women and men in our cohort. In two studies with patients with IIM, female sex was associated with higher pain levels,[8,9] whereas another study reported no significant difference in pain levels between female and male participants.[7]

The prevalence of pain was similar between White and non-Hispanic individuals and others. Race other than White or ethnicity other than Hispanic was also significantly associated with higher pain levels in univariate regression models, but this result lost its significance in multivariable models. These results were somewhat unexpected because prior studies in non-IIM diseases showed enhanced pain sensitivity and disparities in access to care and pain management among Black individuals.[10] However, our results should be interpreted with caution because of small number of individuals with other races in this cohort. Our cohort included only 14 patients who identified themselves as Black and 9 patients who identified themselves as Hispanic. Future studies with IIM cohorts enriched with individuals with other races are warranted to further examine the effect of race on pain levels in patients with IIM.

Pain prevalence was comparable in patients with DM versus non-DM, and IIM subtype was not significantly associated with having higher pain levels. Similar to these results, several studies showed similar pain levels in patients with DM versus non-DM,[9,11] whereas one study showed higher levels of pain in patients with DM compared to patients with IMNM or overlap myositis.[12] Theoretically, patients with DM may experience more pain, as fasciitis is more commonly seen in DM than most other IIM subtypes, which could contribute to pain due to stimulation of nerve endings that are abundant in fascia.[13] The majority of the studies do not support this assumption, which could be related to the problems with the current classification criteria used in IIM. Interestingly, approximately half of the study cohort was taking glucocorticoids despite the long disease duration. This could potentially be due to ongoing pain; however, we could not assess this question in this study.

Although this study was not designed to examine the sources of pain experienced by patients with IIM, it provides important clues into potential causes of pain observed in these patients. Approximately 63% of patients reported muscle pain as well as joint pain. Approximately one-third of patients with IIM had a self-reported diagnosis of fibromyalgia, which may have impacted the prevalence and intensity of pain reported. These rates are comparable to prevalence rates for fibromyalgia in other systemic rheumatic diseases and suggest that similar to patients with other systemic rheumatic diseases, patients with IIM may have abnormalities in central nervous system (CNS) pathways regulating pain.[14] Even though Raynaud phenomenon and other comorbidities such as osteoarthritis may contribute to pain experienced by patients with IIM, no significant differences were observed for the frequencies of osteoarthritis and Raynaud phenomenon in patients who had significant pain versus those who did not.

In the multivariable model, patient global disease activity was the only variable significantly associated with worse pain levels, highlighting the close link between pain and patient-reported disease activity. A prior study showed that one of the leading drivers of discordance between patient and physician global disease activity in IIM was pain.[4] This suggests that, similar to other rheumatic diseases, patients with IIM may often equate pain with disease activity. Studies involving patients with rheumatoid arthritis suggest that abnormalities in CNS regulation of pain, termed central sensitization, may contribute to this discordance.[15] However, our study lacked objective measures of myositis disease activity and central sensitization, which would provide greater clarity on the relationship between inflammatory disease activity, CNS pain regulation, and the patient experience of pain.

Given uncertainty in the underlying pathways linking patient global disease activity and higher pain levels, we conducted a sensitivity analysis excluding patient global activity. In this model, a history of smoking was associated with two times increased risk of pain levels >3 in patients with IIM, after accounting for several sociodemographic variables and comorbidities. The relationship between smoking and pain is challenging to disentangle because it is difficult to discern if the patients who smoke experience heightened pain sensitivity and report more pain or if patients with pain react by smoking more. Prior studies suggest that this relationship may be bidirectional. Pain could increase smoking behavior because of its use as a coping strategy and the small acute analgesic effect of nicotine.[16] Smoking can also enhance pain with loss of the analgesic effect with prolonged exposure and heightened pain in the absence of nicotine.[17] This finding should be further studied in a prospective cohort because it allows for better examination of longitudinal associations between risk factors and outcomes.

Strengths of this study are the following: (1) it is the largest study to date investigating pain prevalence in patients with IIM, and (2) it is the first study to examine non–disease-related factors associated with pain intensity in patients with IIM. However, limitations are also present. First, the cohort predominantly included White and non-Hispanic individuals who agreed to complete semiannual questionnaires, which may not represent the myositis population at large, limiting the generalizability of our findings. Second, the database only included patient-reported variables, relying on patient's knowledge, which may introduce subjectivity and recall bias, and lacked other important variables in myositis assessment, such as muscle enzymes, muscle strength, and physician-reported disease activity results, as well as other medications that can be used in pain management, such as gabapentin. Third, the cross-sectional design limits our understanding of the directionality of results. Although we hypothesized that some factors would contribute to pain while others would be outcomes associated with pain, relationships are likely bidirectional. Lastly, the patients with physician diagnosis of IIM were included in the study; given that the inclusion was not based on classification criteria, this may show variability between sites. Future longitudinal studies involving targeted interventions will be crucial to understanding causal relationships.

This large observational cohort study demonstrated a high prevalence of pain among patients with IIM and identified associations between patient global disease activity, smoking, and significant pain intensity in these patients. Pain was associated with worse fatigue, functional disability, and lower health satisfaction in patients with IIM. These results highlight the critical need to better understand the pain experienced by patients with IIM to best address the needs of the patients and provide patient-centric care.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Leclair V, Tsui H, Hudson M. Pain in autoimmune inflammatory myopathies: a scoping review. RMD Open 2023;9(1):e002591.
2. Saygin D, Alexanderson H, DiRenzo D, et al. The impact of pain on daily activities in patients with idiopathic inflammatory myopathies: report from the OMERACT myositis working group. Semin Arthritis Rheum 2024;67:152476.
3. Mecoli CA, Park JK, Alexanderson H, et al. Perceptions of patients, caregivers, and healthcare providers of idiopathic inflammatory myopathies: an international OMERACT study. J Rheumatol 2019;46(1):106–111.
4. Keret S, Saygin D, Moghadam-Kia S, et al. Discordance between patient- and physician-reported disease activity in adult idiopathic inflammatory myopathy. Rheumatology (Oxford) 2023;62(12):3957–3961.
5. Wolfe F, Michaud K. The National Data Bank for rheumatic diseases: a multi-registry rheumatic disease data bank. Rheumatology (Oxford) 2011;50(1):16–24.
6. Jensen MP, Chen C, Brugger AM. Interpretation of visual analog scale ratings and change scores: a reanalysis of two clinical trials of postoperative pain. J Pain 2003;4(7):407–414.
7. Chandrasekhara Pillai A, Bijoy George T, Dianxu R, et al. Pain is common in myositis and associated with disease activity. Rheumatology (Oxford) 2025;64(2):780–788.
8. Shinjo SK, Kim M, Hoff LS, et al. Pain in individuals with idiopathic inflammatory myopathies, other systemic autoimmune rheumatic diseases, and without rheumatic diseases: a report from the COVAD study. Int J Rheum Dis 2023;26(4):727–739.
9. Ponyi A, Borgulya G, Constantin T, et al. Functional outcome and quality of life in adult patients with idiopathic inflammatory myositis. Rheumatology (Oxford) 2005;44(1):83–88.
10. Rahim-Williams FB, Riley JL III, Herrera D, et al. Ethnic identity predicts experimental pain sensitivity in African Americans and Hispanics. Pain 2007;129(1–2):177–184.
11. Chung YL, Mitchell HL, Houssien DA, et al. A comparative study of outcome in myositis and other musculoskeletal disorders assessed using the Nottingham Health Profile. Clin Exp Rheumatol 2001;19(4):447–450.
12. Landon-Cardinal O, Bachasson D, Guillaume-Jugnot P, et al. Relationship between change in physical activity and in clinical status in patients with idiopathic inflammatory myopathy: a prospective cohort study. Semin Arthritis Rheum 2020;50(5):1140–1149.
13. Noda K, Yoshida K, Ukichi T, et al. Myalgia in patients with dermatomyositis and polymyositis is attributable to fasciitis rather than myositis: a retrospective study of 32 patients who underwent histopathological examinations. J Rheumatol 2017;44(4):482–487.
14. Haliloglu S, Carlioglu A, Akdeniz D, et al. Fibromyalgia in patients with other rheumatic diseases: prevalence and relationship with disease activity. Rheumatol Int 2014;34(9):1275–1280.
15. Lee YC, Bingham CO III, Edwards RR, et al. Association between pain sensitization and disease activity in patients with rheumatoid arthritis: a cross-sectional study. Arthritis Care Res (Hoboken) 2018;70(2):197–204.

16. Patterson AL, Gritzner S, Resnick MP, et al. Smoking cigarettes as a coping strategy for chronic pain is associated with greater pain intensity and poorer pain-related function. J Pain 2012; 13(3):285–292.

17. Carstens E, Anderson KA, Simons CT, et al. Analgesia induced by chronic nicotine infusion in rats: differences by gender and pain test. Psychopharmacology (Berl) 2001;157(1): 40–45.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 141–153
DOI 10.1002/acr.25665

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

Racial and Ethnic Disparities in the Incidence and Prevalence of Low Back Pain in the United States: A Systematic Review

Colleen A. Burke,[1] Rebecca Fillipo,[1] Meira Epplein,[1,2] M. Alan Brookhart,[1] Hayden B. Bosworth,[1,3] and Adam P. Goode[1,4]

Objective. This systematic review synthesizes existing evidence to quantify racial and ethnic disparities in low back pain (LBP) incidence and prevalence in the United States across stages of chronicity (acute, subacute, and chronic LBP).

Methods. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we systematically searched MEDLINE, Embase, the Cumulative Index to Nursing and Allied Health Literature (CINAHL), and Web of Science (through January 7, 2025) for studies reporting LBP incidence or prevalence by race and ethnicity in US adults. Risk of bias was assessed using the Risk of Bias in Non-randomized Studies - of Exposure (ROBINS-E) tool.

Results. Of 8,145 citations, 23 studies met inclusion criteria (10 on incidence, 13 on prevalence). Some incidence studies found higher risk of chronic LBP among Black adults compared to White adults, whereas data on Hispanic and Latino adults remain limited. Prevalence studies showed higher rates in White and American Indian and Alaska Native adults, with lower prevalence in Black, Hispanic and Latino, and Asian adults. Military studies consistently reported that Black service members experienced higher LBP incidence compared to other races. No studies examined the subacute state.

Conclusion. This review highlights persistent race-based differences in LBP, with critical gaps in research on acute LBP incidence and community-based prevalence. Future studies should prioritize population-based research to better capture racial differences in LBP burden and inform targeted interventions.

INTRODUCTION

Nonspecific low back pain (LBP) is pain in the lower back without an identifiable underlying condition. It is often classified by duration: acute (<4 weeks), subacute (4–12 weeks), and chronic (>12 weeks), which helps guide treatment and prognosis. LBP is a leading cause of disability worldwide.[1,2] In 2020, an estimated 619 million people worldwide experienced LBP, with the number of cases expected to reach 843 million by 2050.[3] The estimated lifetime prevalence of LBP has been reported as high as 84%.[4] In the United States, it is the fifth most common reason for all physical visits.[5,6] As the population ages, especially in the United States, the prevalence of LBP, and its associated costs, are expected to continue rising.[7] This imposes a substantial burden on individuals, communities, health care systems, and economies,[5,8] because LBP contributes to reduced quality of life,[9] work absenteeism,[10] and increased health care use.[11,12]

The burden of LBP is not evenly distributed across populations, with social determinants of health (SDOH) playing a critical role in shaping disparities in the incidence, prevalence, and associated health care burden of LBP. SDOH, such as discrimination, socioeconomic status, access to health care, occupational exposures, and neighborhood environments, often contribute to differential LBP risk and experience across racial and ethnic groups.[13–15] Evidence suggests that individuals from historically marginalized communities face higher rates of job insecurity,

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (1F31-AR-084911-01 to Dr Burke).

[1]Duke University School of Medicine, Durham, North Carolina; [2]Duke Cancer Institute, Durham, North Carolina; [3]Durham Veterans Affairs Center of Innovation to Accelerate Discovery and Practice Transformation, Durham, North Carolina; [4]Duke Clinical Research Institute, Duke University, Durham, North Carolina.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25665).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25665.

Address correspondence via email to Colleen A. Burke, PT, DPT, at colleen.a.burke@duke.edu.

Submitted for publication April 1, 2025; accepted in revised form September 25, 2025.

SIGNIFICANCE & INNOVATIONS

- This review examined racial differences in low back pain (LBP) incidence and prevalence across studies conducted over different time periods (1973–2025), finding heterogeneous results that may reflect variation in study populations, methodologies, or changes in social, environmental, or health care factors over time.
- Unlike prior reviews that focus primarily on chronic LBP intensity or treatment approaches, this review reports on various chronicity states, highlighting racial differences in the transition from acute to chronic LBP, which may have implications for long-term disability and pain management.
- By distinguishing among findings from cohort studies, cross-sectional surveys, and administrative claims data, this review provides a nuanced understanding of how racial disparities manifest in different research contexts.
- Findings from this review can inform targeted interventions to reduce disparities in LBP, shape future epidemiologic research, and guide health care policy efforts aimed at addressing inequities in musculoskeletal health.

limited access to timely and high-quality medical care, and greater exposure to psychosocial stressors, all of which have been linked to an increased risk of many health conditions including LBP.[15–17] Limited access to quality health care, including pain management specialists, can contribute to disparities in treatment and pain management for marginalized groups experiencing LBP.[3,18,19] Experiences of discrimination and chronic stress then further exacerbate pain perception and contribute to poorer pain management outcomes for minority populations, which may prolong symptom burden.[20] Research also indicates that Black and Hispanic and Latino populations report greater pain intensity compared to White populations.[21] These differences point to the large contribution that SDOH and disparities play in LBP experiences and burden. A recent systematic review examined SDOH in chronic LBP frequency and severity, reporting that sex, race, ethnicity, education, occupation, and socioeconomic status were key risk factors associated with pain severity, disability, work absenteeism, and opioid use and prescription.[14] Although this review provided invaluable insights into associations between SDOH and specific LBP outcomes, it did not examine how these factors influenced the incidence or the prevalence of LBP, and furthermore, the review was limited to chronic LBP. This leaves a gap in our understanding one of the most influential social drivers of health inequities and disease burden, racial disparities in occurrence of LBP.

These disparities not only contribute to disproportionate suffering and disability, but they also point to a complex interplay of psychosocial, health care, policy, and systemic factors, including differences in access to care, insurance coverage, and provider biases in pain assessment and treatment. These differences result in delayed care and increased emergency department use and health care burden.[22,23] Given these inequities and the expected rise in LBP, a comprehensive systematic review is warranted to build on our knowledge of SDOH and LBP to synthesize existing evidence and quantify racial and ethnic disparities in LBP incidence and prevalence in the United States. This review aims to identify patterns of disparities in the epidemiology of LBP across the stages of chronicity (ie, acute LBP, subacute LBP, and chronic LBP). These findings will be key for informing future studies and public health interventions because a clearer understanding of these disparities can help guide policy efforts to promote equitable pain management strategies and reduce the disproportionate burden of LBP among racially and ethnically minoritized populations. Such knowledge is a critical step toward developing targeted public health strategies, informing policy initiatives, and guiding future research to address disparities and improve LBP-related outcomes for all at risk populations.

METHODS

Study design. A systematic review and meta-analysis was conducted following the 27-item Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement for Reporting Systematic Reviews. The study was exempt from institutional review board review at Duke University. The preregistered protocol for this review is available in the International Prospective Register of Systematic Reviews (PROSPERO; https://www.crd.york.ac.uk/prospero/; registration number CRD42023440302).

Search strategy. MEDLINE (via PubMed), Embase (via Elsevier), Cumulative Index to Nursing and Allied Health Literature (CINAHL; via EBSCOhost), and Web of Science (via Clarivate) databases were searched on June 30, 2023, and repeated on January 7, 2025, using the following search concepts: LBP, race or ethnicity, and incidence or prevalence. To ensure completeness, search strategies were developed in consultation with an expert medical librarian. The full list of search strings for each database can be found in Supplementary Materials 1–4. Filters were applied to include studies involving human participants and published in English. No restrictions were placed on the publication date, allowing for the inclusion of all studies available up to January 2025.

Eligibility criteria. Any study that reported on the incidence or prevalence of LBP by race or ethnicity was considered for inclusion in this study. Study participants were required to be 18 years of age or older. We isolated our search to within the United States because racial disparities are unique to countries and/or regions. Because the goal was to examine nonspecific

LBP, studies investigating a specific etiology, such as trauma, surgery, inflammatory conditions, and pregnancy, were excluded. Pain had to be specified to be in the lower back region; studies that identified back pain alone without reference to the location of the lower back were excluded. Studies needed to provide outcomes of prevalence and/or incidence of LBP, either alone or with some other epidemiologic measures (eg, risk factors, burden of disease). Incidence of LBP was defined as the occurrence of new cases within a specified population over a given period. Studies were permitted to define incidence in various ways, such as first-ever episode of LBP in an individual's lifetime, new episode of LBP after a specified pain-free period, LBP meeting a specific pain threshold (eg, ≥4 of 10 on a pain scale), or LBP that leads to health care use. Prevalence of LBP was defined as the proportion of individuals experiencing LBP at a given time or over a specified period. Prevalence estimates could include point prevalence, period prevalence, or lifetime prevalence. These outcomes were then required to be examined by race and/or ethnicity. Abstracts, narrative review papers, commentaries, pilot studies, case series, single case reports, and qualitative studies were excluded. We excluded studies that recruited only individuals with LBP (eg, treatment-seeking, clinical, or surgical samples) because these samples inherently reflect 100% prevalence and cannot be used to estimate incidence or prevalence within a defined population. Our comprehensive search strategy (Supplementary Materials 1–4) included broad terms related to LBP, epidemiology, and race and ethnicity, designed to maximize sensitivity.

Study selection. Two reviewers (CAB and RF) independently evaluated potentially eligible studies identified through the search strategy detailed in the Search Strategy section. Articles were first screened for eligibility based on review of the title and abstract only, and disagreements were resolved by consensus. Once all articles received two votes for inclusion or exclusion, the two reviewers completed a full-text review to vote for final inclusion or exclusion. All full-text reviews had to receive two votes to be included in the final inclusion list. Consensus for inclusion was obtained with the help of a third reviewer if necessary. At the full-text review stage, we documented one primary reason for article exclusion based on the Patient or Problem, Intervention, Comparison, and Outcome (PICO) framework, prioritizing the first applicable criterion. Previous systematic reviews on this, or a related topic, were hand searched for potential inclusion.[2,14,24] All references were managed and screened on Covidence, a web-based tool for systematic reviews.[25]

Data collection and analysis. Data extraction and study-quality assessment were done by two independent researchers (CAB and RF). Cases of disagreement were solved by a discussion between two reviewers or a third reviewer if necessary. For data extraction, a form was designed in Covidence that included the following variables: author names, location, data source, sample size, year of publication, date of data collection, prevalence and/or incidence (incidence rate, odds ratio, risk ratio, prevalence, etc) of LBP by race and/or ethnicity and associated confidence intervals, and studies' quality score. Age of population, sex (male or female), follow-up period, and other relevant sociodemographic or psychosocial characteristics were also extracted.

Quantitative data synthesis and meta-analysis were not feasible due to heterogeneity in study populations, outcomes, analytical methods, and publication timeframe. Instead, we summarized key descriptive study-level information and results, synthesizing the findings narratively. We grouped findings first based on incidence versus prevalence, followed by general US population versus military personnel, then based on LBP chronicity, and finally chronologically by the date of data collection. Where possible, incidence rates, prevalence estimates, or other epidemiologic measures (ie, odds ratios, risk ratios, and regression coefficients) were extracted.

Quality assessment. Two reviewers (CAB and RF) assessed risk of bias for each included study, using the Risk of Bias in Non-Randomized Studies - of Exposure (ROBINS-E) tool.[26] A third reviewer was used for any conflicts of interest. The ROBINS-E tool offers a systematic method for evaluating the risk of bias in observational epidemiologic studies. It encompasses seven bias domains, each assessed through a set of signaling questions designed to collect key information about the study and its analysis. These domains capture risk of bias due to confounding, measurement of exposure, selection of participants into the study (or into the analysis), postexposure interventions, missing data, measurement of outcome, and selection of the reported result. Finally, an overall judgment was made based on three key considerations. Discrepancies in the risk of bias assessment were resolved through consensus.

RESULTS

Search results and study characteristics. A comprehensive search was conducted across electronic databases, yielding a total of 8,145 citations. After removing duplicate records, 4,999 unique studies remained. Following a title and abstract screening, 4,740 studies were excluded, leaving 259 for full-text evaluation. Upon full-text review, 236 studies were further excluded based on LBP-only samples (no prevalence or incidence estimation possible), pain or LBP etiology, outcomes not separated by race, outcomes other than prevalence or incidence, and improper study design. Ultimately, 23 studies met all inclusion criteria and were incorporated into the final analysis (Figure 1), 10 reported LBP incidence, and 13 reported prevalence. By LBP chronicity, 4 examined acute, 8 examined chronic (including 1 high-impact chronic and 3 transition to chronic

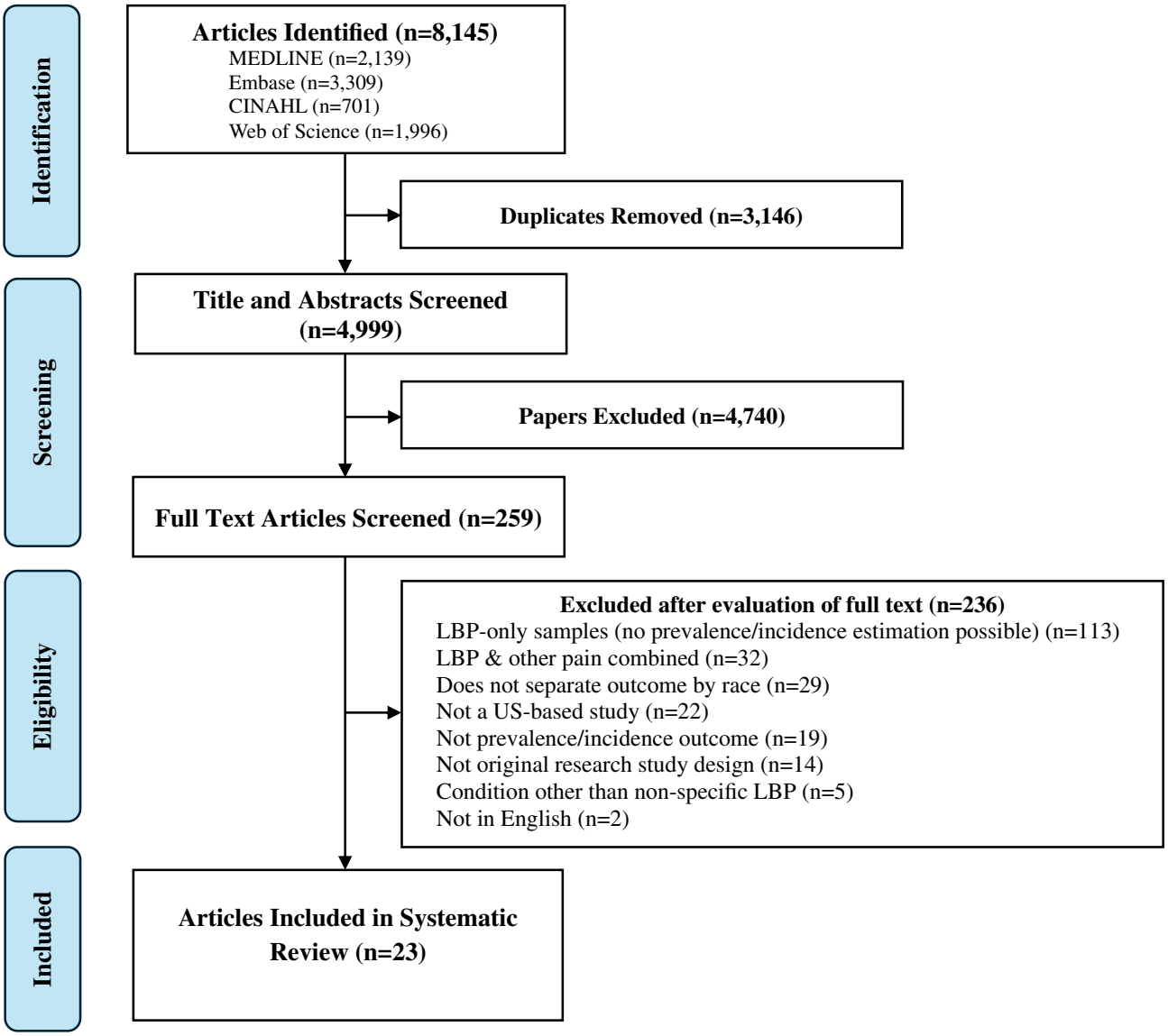


Figure 1. Flow diagram of retrieved, screened, and included studies. CINAHL, Cumulative Index to Nursing and Allied Health Literature; LBP, low back pain.

LBP), and 11 examined general LBP (no chronicity stated). Although the transition to chronic LBP covers the subacute state, no studies specifically examined prevalence or incidence of subacute LBP. Seventeen studies investigated the general US population, whereas six investigated military service members. Sample sizes varied widely, from 1,135 to 1.47 million for prevalence studies and up to 1.48 billion person-years for incidence studies. Some studies focused on specific subpopulations, such as young women or older adults in the general population and vehicle operators or individuals with lumbar radiculopathy in military samples. Racial comparisons varied: four studies compared only Black and White individuals, whereas one included seven racial groups. Data collection methods included medical records, surveys, and interviews, spanning from 1973 to 2025 (Table 1).

LBP incidence. Five studies[21,27–30] examined the incidence of LBP in the general US population (Table 2). Only one study evaluated acute LBP incidence and found that both White and Black adults were over six times more likely than Asian adults to seek treatment for acute LBP over five years.[27] Four studies examined chronic LBP incidence using varying measures: incidence rates, odds ratios, and risk ratios.[21,28–30] Two studies found no racial differences in chronic LBP incidence, one using inverse probability weighting–adjusted odds ratios and the other using raw incidence counts in young women.[28,29] Two others found increased incidence for Black adults compared to White adults.[21,30] Overall, half of the studies reported significant racial differences in LBP incidence or prevalence, with all significant results showing greater burden in Black adults.

Table 1. Study characteristics*

Author, publication year	Region	Date of data collection	Population site/ data source	Study size	Female sex, %	Age, y (%)	Chronicity
LBP incidence studies							
Waterman, 2012	National	2004–2008	NEISS	65.63 million (1.48 billion person-years)	48.5	<20 (8.9) 20–39 (46.5) 40–59 (33.2) 60–79 (8.2) 80+ (3.2)	Acute
Plesh, 2012[a]	National	1987–1992	NGHS (California, Ohio, and the District of Columbia)	732	100	19–26	Chronic
Stevans, 2021	National	2016–2019	TARGET Trial; 4 health systems (Pittsburgh, PA; Boston, MA; Salt Lake City, UT; & Baltimore, MD)	5,233	58	18–40 (31) 41–60 (39) 61+ (30)	Chronic
Roseen, 2023	National	2016–2019	TARGET Trial; 4 health systems (Pittsburgh, PA; Boston, MA; Salt Lake City, UT; & Baltimore, MD)	9,088	59	Mean 49.9	Chronic & high-impact chronic
Burke, 2025	North Carolina	2022–2023	BACk Study; community-based cohort	131	58.8	Mean 56.9 (SD 13.7)	Chronic
Ernat, 2012	Military service members	1998–2006	DMED	791,526 person-years	NR	NR	Acute
Knox, 2012	Military service members	2000–2009	DMSS	12,399,276 person-years	NR	<20 (10.8) 20–29 (55.1) 30–39 (25.7) 40+ (8.3)	Acute
Knox, 2011	Military service members	1998–2006	DMED	13,754,261 person-years	14.3	<20 (8.1) 20–29 (53.6) 30–39 (28.3) 40+ (10.2)	General
Knox, 2014[a]	Military vehicle operators	1998–2006	DMED	213,024	12.6	<20 (10.0) 20–29 (58.3) 30–39 (25.0) 40+ (6.8)	General
Schoenfeld, 2012[a]	Military service members	2000–2009	DMED	13,813,333 person-years	14.6	<20 (7.9) 20–24 (33.4) 25–29 (21.1) 30–24 (14.5) 35–39 (12.8) 40+ (10.3)	General (lumbar radiculopathy)

(Continued)

Table 1. (*Cont'd*)

Author, publication year	Region	Date of data collection	Population site/ data source	Study size	Female sex, %	Age, y (%)	Chronicity
LBP prevalence studies							
Carey, 1996	North Carolina	1992	North Carolina; UNC-CH Survey Research Unit (care seeking)	8,067	NR	NR	Acute (severe)
Hardt, 2008	National	1992–2002	NHANES	10,271	53.3	Mean 50.2	Chronic
Freburger, 2009	North Carolina	1992 & 2006	North Carolina; UNC-CH Survey Research Unit	1992: 8,067 2006: 9,924	NR	NR	Chronic
Carey, 2010	North Carolina	2006	North Carolina; UNC-CH Survey Research Unit	9,924	61.4	Mean 52.5	Chronic
Hogans, 2021[a]	National: adults 65+	2017	CMS	1,477,594	57.7	Mean 75.4	Chronic
Nagi, 1973	Columbus, OH	NR	SMSA	1,135	55.6	18–34 (42.2) 35–49 (28.4) 50–64 (25.3)	General
Deyo, 1987	National	1976–1980	NHANES II	10,404	NR	NR	General
Andersen, 2003[a]	National: adults 60+	1988–1994	NHANES III	5,724	NR	NR	General
Deyo, 2006	National	2002	NHIS	31,044	NR	NR	General
Plesh, 2011	National	2000–2005	NHIS	189,992	52	NR	General
Bauer, 2012	National: White, Latino, and Asian American	2002–2003	NLAAS	4,864	NR	NR	General
Peng, 2018	National	2015	NHIS	32,060	54.4	Mean 47.0 (SD 0.2)	General
Mullinax, 2023	Department of Navy	2009–2015 (annual estimates)	MDR	Low: 424,460 (FY 2014) High: 437,053 (FY 2011)	15.12 (FY 2015)	FY 2015: <25 (46.13) 25–34 (36.20) 35+ (17.67)	General

* General includes acute, subacute, and chronic states of LBP. CMS, Centers for Medicare & Medicaid Services; DMED, Defense Medical Epidemiology Database; DMSS, Defense Medical Surveillance System; FY, fiscal year; LBP, low back pain; MDR, Military Health System Data Repository; NEISS, National Electronic Injury Surveillance System; NGHS, National Heart, Lung and Blood Institute Growth and Health Study; NHANES, National Health and Nutrition Examination Survey; NHIS, National Health Interview Survey; NLAAS, National Latino and Asian American Study; NR, not reported; SMSA, standard metropolitan statistical area; UNC-CH, University of North Carolina at Chapel Hill.
[a] Special population study.

Table 2. Incidence estimates of racial and ethnic disparities in LBP*

Study, year	LBP chronicity	Date of data collection	Incidence measure	Incidence estimate					
				White	Black	Latino	AAPI	AI/AN	Other
General US population									
Waterman, 2012	Acute	2004–2008	IRR per 1,000 person-years	1.23[a]	1.38[a]	0.4	0.2	2.1	–
Plesh, 2012[b]	Chronic	1987–1992	Incidence per 1,000 person-years	114	112	–	–	–	–
Stevans, 2021	Chronic	2016–2019	IPW-adjusted OR	1	1.22	–	–	–	1.18
Roseen, 2023	Chronic	2016–2019	IPW-adjusted OR	1	1.49[c]	1.26	–	–	–
Roseen, 2023	High-impact chronic	2016–2019	IPW-adjusted OR	1	1.90[c]	1.83[c]	–	–	–
Burke, 2025	Chronic	2022–2023	RR	1	1.76[c]	–	–	–	–
Military specific populations									
Ernat, 2012[d]	Acute	1998–2006	IR per 1,000 person-years	32.2	42.2	–	–	–	30.3
Knox, 2012	Acute	2000–2009	IR per 1,000 person-years	37.1	43.7[c]	35.9[c]	30.7[c]	32.4[c]	–
Knox, 2011	General	1998–2006	IR per 1,000 person-years	38.9[e]	47.4[e]	–	–	–	38.5
Knox, 2014[b]	General	1998–2006	IR per 1,000 person-years	53.5[e]	59.0[e]	–	–	–	45.4
Schoenfeld, 2012[b]	General (lumbar radiculopathy)	2000–2009	IR per 1,000 person-years	5.02[f]	4.64	–	–	–	4.26[f]

* Latino captures Hispanic, Latino, and Mexican-American. AAPI, Asian American and Pacific Islander; AI/AN, American Indian and Alaska Native and Native American; IPW, inverse probability weight; IR, incidence rate; IRR, incidence rate ratio; LBP, low back pain; OR, odds ratio; RR, risk ratio.
[a] Statistically significant difference with AAPI as the referent group.
[b] Special population study.
[c] Statistically significant difference with White as the referent group.
[d] No statistical significance measure or interpretation provided.
[e] Statistically significant difference with other as the referent group.
[f] Statistically significant difference with Black as the referent group.

Five studies[31–35] examined racial disparities in LBP incidence among military service members (Table 2). Two reported on acute LBP[31,33] and three on general LBP[32,34,35] (one of which examined lumbar radiculopathy[35]). All reported unadjusted incidence per 1,000 person-years. Race was a statistically significant factor in four studies.[32–35] In three studies, Black service members consistently experienced higher incidence than White service members, and other racial groups reported lower incidence estimates.[32–34] The fourth study did not report any statistical significance measures, but the reported incidence rate for Black service members was higher compared to White service members (incidence rate = 42.2 and 32.2 per 1,000 person-years, respectively)[31]. The fifth study on lumbar radiculopathy reported that White service members had higher incidence than Black service members (referent), whereas all other racial groups had lower rates.[35]

LBP Prevalence. Twelve studies[36,37,38–46] examined LBP prevalence in the US general population (Table 3). Only one study evaluated acute LBP prevalence, using 1992 data from North Carolina.[38] It found higher rates of treatment for severe acute LBP among White adults compared to Black adults. Four studies assessed chronic LBP prevalence using data from 1992 to 2017[6,40–42]. One national study reported higher chronic LBP prevalence in White compared to Hispanic and Latino adults.[39] Two North Carolina–based studies found no racial differences in 1992 but significant disparities by 2006, with higher prevalence in White adults compared to Latino adults.[40,41] A study using Centers for Medicare & Medicaid Services (CMS) data found the highest chronic LBP prevalence among White older adults, followed by American Indian and Alaska Native adults, with the lowest prevalence in the "other race" category.[42] Seven additional studies examined general LBP prevalence in the US population.[36,37,43–46] One was regional;[36] six used national data.[37,43–46] Among all 12 general population prevalence studies, 7 reported statistically significant racial disparities, all showing higher LBP prevalence in White adults.[38–41,43,45,46] However, three studies did not report statistical significance.[36,37,42] Early data from 1973 showed minimal difference between Black and White adults, although significance was not tested.[36] Two studies by Deyo et al[43,45] found significantly lower LBP prevalence among Black and Asian adults compared to White and American Indian and Alaska Native adults[43,45].

Only one study investigated the prevalence of LBP in the military (Table 3).[47] Mullinax et al used the Military Health System Data Repository from 2009 to 2015 and reported annual prevalences of general LBP in active Navy and Marine personnel. This study compared White service members to all other racial groups (combined) and found that for four of the seven years, all other racial groups had a small but statistically significant higher risk of LBP compared to White service members.[47]

Trends in LBP incidence and prevalence over time among the general US population. Only one study assessed the incidence of acute[27] and high-impact chronic LBP,[21] limiting our ability to evaluate temporal trends in incidence for these

Table 3. Prevalence estimates of racial and ethnic disparities in LBP*

Study, year	LBP chronicity	Date of data collection	Prevalence estimate								
			White	Black	Latino	AI/AN	Asian	NH/PI	Other	Mult.	Unkn.
General US population											
Carey, 1996	Acute (severe)	1992	8.0	5.0[a]	–	–	–	–	–	–	–
Hardt, 2008	Chronic	1999–2002	10.8	9.0	5.4[a]	–	–	–	9.2	–	–
Freburger, 2009	Chronic	1992	4.1	3.0	–	–	–	–	4.1	–	–
Freburger, 2009	Chronic	2006	10.5	9.8	6.3[a]	–	–	–	9.1	–	–
Carey, 2010	Chronic	2006	10.4	9.8	6.3[a]	–	–	–	–	–	–
Hogans, 2021[b,c]	Chronic	2017	15.1	12.7	13.4	14.9	14.4	–	11.9	–	12.8
Nagi, 1973[c]	General	NR	17.9	18.6	–	–	–	–	–	–	–
Deyo, 1987	General	1976–1980	14.2	13.4[a]	–	–	–	–	9.3[a]	–	–
Andersen, 2003[b]	General	1988–1994	21.2	23.5	21.8	–	–	–	–	–	–
Deyo, 2006	General	2002	27.4	23.9[a]	24.3	35.0[a]	19.0[a]	26.8	–	–	–
Plesh, 2011	General	2000–2005	29.5	24.6[a]	24.4[a]	–	–	–	–	–	–
Bauer, 2012[b]	General	2002–2003	14.0	–	13.7	–	9.3[a]	–	–	–	–
Peng, 2018[c]	General	2015	31.8	28.4	26.8	37.7	19.0	–	–	34.6	–

Military specific populations			Ratio of prevalence among all other races to White adults	95% CI
Mullinax, 2023	General	2009	1.04[a]	1.02–1.06
		2010	1.02	1.00–1.04
		2011	1.02	1.00–1.04
		2012	1.03[a]	1.01–1.05
		2013	1.02	1.00–1.03
		2014	1.02[a]	1.01–1.04
		2015	1.03[a]	1.02–1.05

* Latino captures Hispanic, Latino, and Mexican-American. Prevalence ratios for military specific populations reflect prevalence comparisons between White adults and all other races, with White as the referent group. AI/AN, American Indian and Alaska Native; CI, confidence interval; LBP, low back pain; NH/PI, Native Hawaiian and Pacific Islander; NR, not reported.
[a] Statistically significant difference with White as the referent group.
[b] Special population study.
[c] No statistical significance measure or interpretation provided.

conditions. Furthermore, due to methodologic heterogeneity of studies, only descriptive incidence estimates were presented across studies by race and ethnicity; these data are not intended to reflect or imply longitudinal trends in disparities. Among the four studies reporting chronic LBP incidence,[21,28–30] one study focused on a special population[28] and two studies used the same data source.[21,29] Estimates include incidence rates,[27] raw cumulative incidence[28], odds ratios,[21,29] and risk ratios.[30] A comparison between Stevans et al and Burke et al displays an increase in the observed disparity in chronic LBP incidence between White and Black adults (Figure 2A). However, it is important to note that Burke et al examined data exclusively from non–care-seeking individuals in two communities in North Carolina and reported risk ratios. In contrast, Stevans et al used a more geographically diverse sample of health care–seeking adults and reported odds ratios, which may overestimate risk when the outcome is common. Although these studies both examine incident chronic LBP in recent years, these marked differences in study design make a direct comparison between the two study findings challenging. For Plesh,[48] an incidence rate was calculated using reported data on raw cumulative incidence. This variability reflects differences in study design and outcome reporting, and estimates are presented as originally reported or derived using available data.

Similar to incidence, only one study examined the prevalence of acute LBP, making temporal trends unobservable.[38] For chronic LBP, prevalence data were available from studies published between 1992 and 2021.[39–42] These data indicate an overall increase in prevalence from 3.9% to 14.8%. Across studies examining chronic LBP, White adults consistently exhibited higher prevalence rates than Black adults. However, the magnitude of this disparity did not appear to increase over time (Figure 2B). A similar trend of increasing overall prevalence over time was observed for general LBP. The earliest study published in 1973 reported a prevalence of 17.9%,[36] compared to 31.4% in 2018.[37] In contrast to chronic LBP, racial disparities in general LBP prevalence varied across studies, with different racial groups experiencing higher prevalence in different studies.

Risk of bias assessment and heterogeneity. Quality assessment using the ROBINS-E tool demonstrated a wide range of risk of bias scores. Overall, 39.1% of studies were classified as a low risk of bias (n = 9), 47.8% as a moderate risk of bias (n = 11), and 13.0% as a high risk of bias (n = 3) (Table 4). Of the included studies, most risk of bias concerns came from unobserved confounding, selection of participants, missing data, and measurement of the outcome.

A meta-analysis and formal tests for heterogeneity (eg, Cochran's Q, I^2 statistic) were not conducted due to substantial clinical and methodologic heterogeneity across studies.

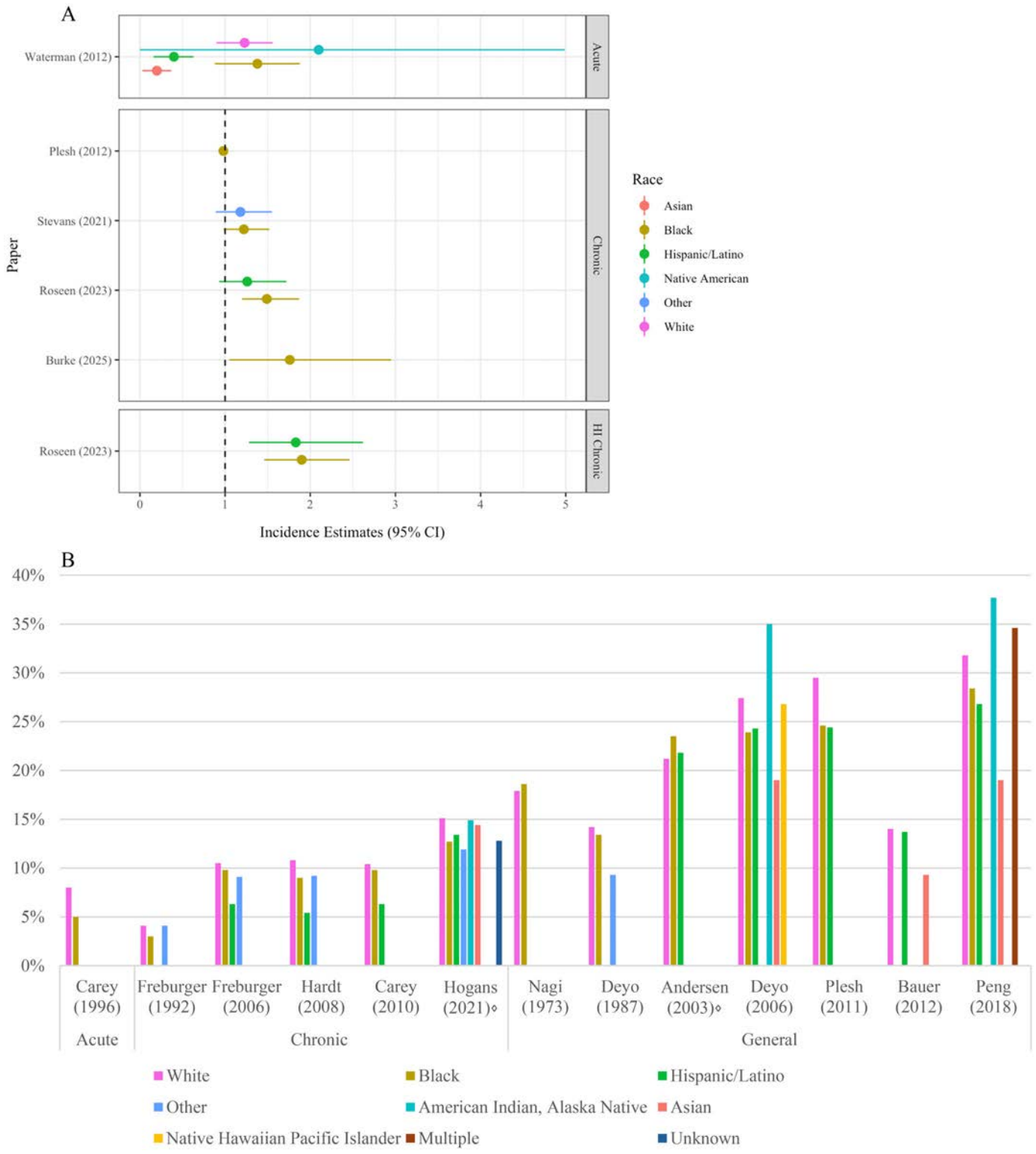


Figure 2. Racial and ethnic LBP disparities over time among the general US population in (A) incidence and (B) prevalence. For A, Incidence estimates reflect the measure reported in each study: incidence rates (Waterman and Plesh), odds ratios (eg, Stevans and Roseen), and risk ratios (eg, Burke). For Plesh, an incidence rate was calculated using reported data on raw cumulative incidence. The vertical dashed line at 1.0 represents the null value for relative measures but does not apply to raw incidence estimates. For Plesh, Stevans, Roseen, and Burke, incidence estimates are reported as odds or risk ratios with White race as the referent group. The dashed vertical line at 1.0 represents the null value for these relative measures. ⋄Special population study. CI, confidence interval; HI, high impact; LBP, low back pain. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25665/abstract.

Table 4. ROBINS-E risk of bias*

Author, year	Confounding	Measurement of exposure	Selection of participants	Postexposure interventions	Missing data	Measurement of outcome	Selection of reported result	Overall quality
Plesh, 2012[a]	1	1	1	0	0	1	0	Moderate risk
Waterman, 2012	0	0	1	0	0	0	0	Moderate risk
Stevans, 2021	0	0	1	0	1	0	0	Moderate risk
Roseen, 2023	0	0	1	0	1	0	0	Moderate risk
Burke, 2025	0	0	1	0	1	1	0	Moderate risk
Ernat, 2012	0	0	0	0	0	0	0	Low risk
Knox, 2012	0	0	0	0	0	1	0	Moderate risk
Knox, 2011	0	0	0	0	0	0	0	Low risk
Knox, 2014[a]	0	0	0	0	0	0	0	Low risk
Schoenfeld, 2012[a]	0	0	0	0	0	0	0	Low risk
Carey, 1996	1	1	1	0	1	2	1	High risk
Hardt, 2008	0	0	0	0	0	0	0	Low risk
Freburger, 2009	1	0	0	0	1	2	1	High risk
Carey, 2010	1	0	0	0	1	2	0	High risk
Hogans, 2021[a]	1	0	0	0	0	0	0	Moderate risk
Nagi, 1973	1	0	1	0	0	0	1	Moderate risk
Deyo, 1986	1	0	1	0	0	0	1	Moderate risk
Andersen, 2003[a]	1	0	0	0	0	0	1	Moderate risk
Deyo, 2006	1	0	0	0	1	1	1	Low risk
Plesh, 2011	0	0	0	0	0	0	0	Low risk
Bauer, 2012[a]	1	0	1	0	0	0	0	Moderate risk
Peng, 2018	0	0	0	0	0	0	0	Low risk
Mullinax, 2023	0	0	0	0	0	0	0	Low risk

* ROBINS-E; Risk of Bias in Non-randomized Studies-of Exposure.
[a] Special population study.

Specifically, included studies varied widely in outcome measures (eg, incidence rates, odds ratios, risk ratios, prevalence estimates), population characteristics (eg, military vs general US population), definitions of LBP chronicity (acute, chronic, general), and how racial and ethnic categories were operationalized. Many outcomes were reported by a single study or were not comparable across studies, precluding meaningful statistical synthesis. As a result, findings were summarized narratively, and results were grouped by outcome type, chronicity, and population.

DISCUSSION

This systematic review highlights the presence of racial disparities in the incidence and prevalence of LBP in both the US general and military populations. Across studies, several reported higher incidence and lower prevalence of LBP among Black adults compared to White adults, although these patterns were not always uniform. Evidence for other racial and ethnic groups were limited. These differences were most consistently observed in chronic LBP, whereas findings for general LBP were mixed. Heterogeneity in study design, case definitions, and timeframes complicates conclusions about temporal trends or consistency of disparities across populations.

The findings of this review highlight an important consideration: how race is conceptualized and reported and the implications of these decisions. All included studies reported on White participants, and all but one included Black participants. Hispanic and Latino ethnicity was reported in most studies (three incidence studies and nine prevalence studies), followed by "other" race categories in a smaller subset (one incidence study and five prevalence studies). Reporting of other racial groups including Asian, Native Hawaiian and Pacific Islander, American Indian, and Alaska Native participants was less frequent. This underrepresentation makes it challenging to generate stable estimates of incidence and prevalence.

Further complicating interpretation, the way race was categorized varied across studies. For example, one study compared "Black versus White versus other," whereas another distinguished "White versus Black versus Asian versus American Indian versus unknown." Although inconsistent categorizations limit comparability across studies, simply using uniform categories is not a complete solution. Racial and ethnic identity is multifaceted and context dependent, often reflecting structural inequities and lived experiences. Frequently, people identify with more than race or ethnicity, thus introducing further heterogeneity that complicates interpretation. The NIH requires demographic reporting so that research is representative, reproducible, and relevant to all populations and so disparities can be identified and addressed rather than overlooked. However, although baseline demographics may be reported, many epidemiologic studies on LBP incidence and prevalence did not stratify their outcomes by race. As a result, critical opportunities to understand and address disparities are often missed. Recognizing both the limitations of categorical approaches and the importance of transparent, consistent reporting is essential to advancing equity-focused LBP research.

A notable finding is that Black adults were seen to have a higher incidence of LBP but lower prevalence compared to White adults. Several factors may explain this apparent paradox. Incidence estimates were often derived from health care use data, which capture individuals who seek care for a new episode of LBP, typically those with more severe or disruptive symptoms. In contrast, prevalence estimates often came from self-reported surveys in which individuals could report LBP regardless of severity or whether they sought care. If racially minoritized groups are less likely to report milder symptoms in surveys but more likely to experience severe pain that requires medical attention, this may help explain the paradox. Another possibility is that the data accurately reflect real differences: higher incidence with lower prevalence often indicates more short-lived episodes, whereas lower incidence with higher prevalence suggests longer-duration, chronic cases. One possible interpretation is that acute, short-lived but intense episodes may be more common among Black adults, whereas chronic LBP could be more common among White adults. Beyond measurement differences, disparities in pain severity, systemic barriers to health care access, and medical mistrust may also contribute, which may result in Black patients disengaging from follow-up care so their prevalent pain is no longer captured, despite a high burden.[49] Structural inequities in health care access, pain management, and occupational exposures likely influence these disparities.[49,50] However, the evidence base is too limited to draw firm conclusions, and more longitudinal research is necessary.

With regards to health care patterns, this review highlights the need for more community-based and population-level studies on the incidence and prevalence of LBP by race, particularly acute LBP. Research on nonspecific acute LBP is extremely limited. Most studies focus on general or chronic LBP, and the only acute LBP prevalence study examined severe cases resulting in care seeking in 1992, substantially limiting generalizability, especially for racial disparities. Acute LBP may be less commonly studied because of its short duration and high likelihood of resolution, which create methodologic challenges in measuring incidence outside health care settings. Prospective surveillance of short-duration symptoms would require large, healthy cohorts with frequent follow-up, a resource-intensive design feasible mainly in structured systems such as the military. Several studies relied on health care–seeking cohorts, which may not capture true incidence or prevalence estimates because marginalized groups often face barriers accessing care and may not seek treatment for acute episodes.[51] Longitudinal, community- and workplace-based cohorts could provide a more comprehensive picture of acute LBP onset. Furthermore, understanding racial disparities in acute LBP is crucial for determining whether disparities emerge at the onset of LBP or develop as pain persists, which may provide insight into transitions from acute to chronic LBP.

Identifying how and when disparities emerge is essential for uncovering mechanisms and developing timely, targeted interventions to advance equity in prevention and treatment. Some studies suggested higher incidence in racially minoritized groups, particularly Black adults. If these findings are confirmed, early intervention is critical to preventing chronic pain progression. Reports of lower prevalence in these groups may reflect underreporting, differences in symptom perception, or barriers to diagnosis and care rather than a truly lower burden of disease. Incidence data from health care encounters may therefore provide a more accurate indicator of severe or disabling cases requiring clinical attention. Focusing on early identification and treatment could mitigate disparities in long-term outcomes, even when self-reported prevalence appears lower. Inconsistent prevalence patterns across racial groups highlight the need for culturally responsive pain assessment and management. Expanding community-based and workplace interventions and ensuring equitable access to pain management resources are essential to reducing disparities and improving outcomes.

This systematic review is subject to limitations. Variability in study designs, data sources, and LBP case definitions introduces heterogeneity, complicating comparisons. Definitions of acute, chronic, and general LBP were similar but inconsistent, and studies used different racial categorizations, further limiting cross-study comparability. The lack of longitudinal data on acute LBP incidence in the general population is a major gap, with only one relevant study from 2012. More research is needed on acute LBP incidence outside health care settings and the factors driving disparities in the transition to chronic LBP. Similarly, limited research on LBP prevalence in military populations prevents a full understanding of trends and disparities in this setting. Additionally, many studies lacked adjustment for health care access, occupational exposures, and SDOH, factors that likely contribute to LBP disparities. Race is an imperfect proxy for systemic discrimination, and LBP disparities often reflect broader structural inequities in health care use and occupational risk. Most included studies reported unadjusted estimates, increasing the risk of bias. Future research should incorporate these factors to better understand the underlying drivers of disparities.

Despite these limitations, this review provides a comprehensive synthesis of racial disparities in LBP incidence and prevalence in both general and military populations. Addressing key knowledge gaps, including acute LBP incidence in the general population, community-based incidence research, acute LBP prevalence studies, updated chronic LBP prevalence estimates, and evolving military prevalence trends, will be essential for informing targeted interventions and improving pain management strategies across diverse populations.

ACKNOWLEDGMENTS

The authors would like to thank Lesley Skalla from the Duke University Medical Center Library & Archives for her invaluable assistance in developing and refining the systematic search strategy for this review.

Her expertise in literature searching and database navigation greatly contributed to the comprehensiveness and rigor of our study. The development of this systematic review involved minimal use of large language models (LLMs). LLM assistance was limited to grammar and spell checking for clarity and readability. Additionally, LLMs were used for minor refinements in R code related to figure creation, ensuring accuracy in data visualization. All substantive scientific content, data interpretation, and analysis were conducted independently by the authors.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020;8(6):299.
2. Hoy D, Bain C, Williams G, et al. A systematic review of the global prevalence of low back pain. Arthritis Rheum 2012;64(6):2028–2037.
3. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5(6):e316–e329.
4. Balagué F, Mannion AF, Pellisé F, et al. Non-specific low back pain. Lancet 2012;379(9814):482–491.
5. Hoy D, Brooks P, Blyth F, et al. The epidemiology of low back pain. Best Pract Res Clin Rheumatol 2010;24(6):769–781.
6. Hart LG, Deyo RA, Cherkin DC. Physician office visits for low back pain. Frequency, clinical evaluation, and treatment patterns from a U.S. national survey. Spine 1995;20(1):11–19.
7. Manchikanti L, Singh V, Falco FJ, et al. Epidemiology of low back pain in adults. Neuromodulation 2014;17(suppl 2):3–10.
8. Deyo RA, Cherkin D, Conrad D, et al. Cost, controversy, crisis: low back pain and the health of the public. Annu Rev Public Health 1991;12(1):141–156.
9. Agnus Tom A, Rajkumar E, John R, et al. Determinants of quality of life in individuals with chronic low back pain: a systematic review. Health Psychol Behav Med 2022;10(1):124–44.
10. Driscoll T, Jacklyn G, Orchard J, et al. The global burden of occupationally related low back pain: estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014;73(6):975–981.
11. Spears CA, Hodges SE, Kiyani M, et al. Health care resource utilization and management of chronic, refractory low back pain in the United States. Spine 2020;45(20):E1333–E1341.
12. Kirsch EP, Yang LZ, Lee HJ, et al. Healthcare resource utilization for chronic low back pain among high-utilizers. Spine J 2024;24(4):601–616.
13. Mathieu J, Roy K, Robert M, et al. Sociodemographic determinants of health inequities in low back pain: a narrative review. Front Public Health 2024;12:1392074.
14. Karran EL, Grant AR, Moseley GL. Low back pain and the social determinants of health: a systematic review and narrative synthesis. Pain 2020;161(11):2476–2493.
15. Institute of Medicine (US) Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. Smedley BD, Stith AY, Nelson AR, eds. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. National Academies Press (US); 2003.
16. Yang H, Lu ML, Haldeman S, et al. Psychosocial risk factors for low back pain in US workers: data from the 2002-2018 quality of work life survey. Am J Ind Med 2023;66(1):41–53.
17. Khubchandani J, Price JH. Association of job insecurity with health risk factors and poorer health in American workers. J Community Health 2017;42(2):242–251.
18. Lind KE, Flug JA. Sociodemographic variation in the use of conservative therapy before MRI of the lumbar spine for low back pain in the era of public reporting. J Am Coll Radiol 2019;16(4 Pt B):560–569.
19. Gliedt JA, Spector AL, Schneider MJ, et al. Disparities in chiropractic utilization by race, ethnicity and socioeconomic status: a scoping review of the literature. J Integr Med 2023;21(2):159–167.
20. Wilson JM, Booker SQ, Goodin BR, et al. The intersectionality of chronic pain stigma and racial discrimination in Black and White adults with chronic low back pain. Pain Med 2025;26(3):156–165.
21. Roseen EJ, Smith CN, Essien UR, et al. Racial and Ethnic disparities in the incidence of high-impact chronic pain among primary care patients with acute low back pain: a cohort study. Pain Med 2023;24(6):633–643.
22. Nguyen LH, Dawson JE, Brooks M, et al. Disparities in pain management. Anesthesiol Clin 2023;41(2):471–488.
23. Carey TS, Garrett JM. The relation of race to outcomes and the use of health care services for acute low back pain. Spine 2003;28(4):390–394.
24. Meucci RD, Fassa AG, Faria NM. Prevalence of chronic low back pain: systematic review. Rev Saude Publica 2015;49:1.
25. *Covidence*. Version. 2025. Accessed March 27, 2023.
26. Higgins JPT, Morgan RL, Rooney AA, et al. A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E). Environ Int 2024;186:108602.
27. Waterman, B. R., Belmont, P. J., & Schoenfeld, A. J. (2012). Low back pain in the United States: Incidence and risk factors for presentation in the emergency setting. The Spine Journal, 12(1), 63–70. https://doi.org/10.1016/j.spinee.2011.09.002
28. Plesh, O., Gansky, S. A., & Curtis, D. A. (2012). Chronic pain in a biracial cohort of young women. The Open Pain Journal, 5(1), 24–31. https://doi.org/10.2174/1876386301205010024
29. Stevans, J. M., Delitto, A., Khoja, S. S., Patterson, C. G., Smith, C. N., Schneider, M. J., Freburger, J. K., Greco, C. M., Freel, J. A., Sowa, G. A., Wasan, A. D., Brennan, G. P., Hunter, S. J., Minick, K. I., Wegener, S. T., Ephraim, P. L., Friedman, M., Beneciuk, J. M., George, S. Z., & Saper, R. B. (2021). Risk factors associated with transition from acute to chronic low back pain in us patients seeking primary care. JAMA Network Open, 4(2), e2037371. https://doi.org/10.1001/jamanetworkopen.2020.37371
30. Burke, C. A., Fillipo, R., George, S. Z., Kapos, F. P., Kosinski, A. S., Ford, E., Danyluk, S. T., Kingsbury, C. A., Seebeck, K., Lewis, C. E., Plez, C., Brown, M. C., & Goode, A. P. (2025). Transition from acute to chronic low back pain in a community-based cohort. The Journal of Pain, 26, 104704. https://doi.org/10.1016/j.jpain.2024.104704
31. Ernat, J., Knox, J., Orchowski, J., & Owens, B. (2012). Incidence and risk factors for acute low back pain in active duty infantry. Military Medicine, 177(11), 1348–1351. https://doi.org/10.7205/milmed-d-12-00183
32. Knox, J., Orchowski, J., Scher, D. L., Owens, B. D., Burks, R., & Belmont, P. J. (2011). The Incidence of low back pain in active duty United States military service members. Spine, 36(18), 1492–1500. https://doi.org/10.1097/brs.0b013e3181f40ddd

33. Knox, J. B., Orchowski, J. R., & Owens, B. (2012). Racial differences in the incidence of acute low back pain in United States military service members. Spine, 37(19), 1688–1692. https://doi.org/10.1097/brs.0b013e318255a07b
34. Knox, J. B., Orchowski, J. R., Scher, D. L., Owens, B. D., Burks, R., & Belmont, P. J. (2014). Occupational driving as a risk factor for low back pain in active-duty military service members. The Spine Journal, 14(4), 592–597. https://doi.org/10.1016/j.spinee.2013.06.029
35. Schoenfeld, A. J., Laughlin, M., Bader, J. O., & Bono, C. M. (2012). Characterization of the incidence and risk factors for the development of lumbar radiculopathy. Journal of Spinal Disorders & Techniques, 25 (3), 163–167. https://doi.org/10.1097/bsd.0b013e3182146e55
36. Nagi SZ, Riley LE, Newby LG. A social epidemiology of back pain in a general population. J Chronic Dis 1973;26(12):769–779.
37. Peng T, Pérez A, Pettee Gabriel K. The association among overweight, obesity, and low back pain in U.S. adults: a cross-sectional study of the 2015 National Health Interview Survey. J Manipulative Physiol Ther 2018;41(4):294–303.
38. Carey, T. S., Evans, A. T., Hadler, N. M., Lieberman, G., Kalsbeek, W. D., Jackman, A. M., Fryer, J. G., & McNutt, R. A. (1996). Acute severe low back pain. Spine, 21(3), 339–344. https://doi.org/10.1097/00007632-199602010-00018
39. Hardt, J., Jacobsen, C., Goldberg, J., Nickel, R., & Buchwald, D. (2008). Prevalence of chronic pain in a representative sample in the United States. Pain Medicine, 9(7), 803–812. https://doi.org/10.1111/j.1526-4637.2008.00425.x
40. Freburger, J. K., Holmes, G. M., Agans, R. P., Jackman, A. M., Darter, J. D., Wallace, A. S., Castel, L. D., Kalsbeek, W. D., & Carey, T. S. (2009). The rising prevalence of chronic low back pain. Archives of Internal Medicine, 169(3), 251. https://doi.org/10.1001/archinternmed.2008.543
41. Carey, T. S., Freburger, J. K., Holmes, G. M., Jackman, A., Knauer, S., Wallace, A., & Darter, J. (2010). Race, care seeking, and utilization for chronic back and neck pain: Population perspectives. The Journal of Pain, 11(4), 343–350. https://doi.org/10.1016/j.jpain.2009.08.003
42. Hogans, B. B., Siaton, B. C., Taylor, M. N., Katzel, L. I., & Sorkin, J. D. (2021). Low back pain and substance use: Diagnostic and administrative coding for opioid use and dependence increased in U.S. older adults with low back pain. Pain Medicine, 22(4), 836–847. https://doi.org/10.1093/pm/pnaa428
43. Deyo, R. A., & Tsui-Wu, Y.-J. (1987). Descriptive epidemiology of low-back pain and its related medical care in the United States. Spine, 12 (3), 264–268. https://doi.org/10.1097/00007632-198704000-00013
44. Andersen, R. E., Crespo, C. J., Bartlett, S. J., Bathon, J. M., & Fontaine, K. R. (2003). Relationship between body weight gain and significant knee, hip, and back pain in older Americans. Obesity Research, 11(10), 1159–1162. Portico. https://doi.org/10.1038/oby.2003.159
45. Deyo, R. A., Mirza, S. K., & Martin, B. I. (2006). Back pain prevalence and visit rates. Spine, 31(23), 2724–2727. https://doi.org/10.1097/01.brs.0000244618.06877.cd
46. Bauer, A. M., Chen, C.-N., & Alegría, M. (2012). Prevalence of physical symptoms and their association with race/ethnicity and acculturation in the United States. General Hospital Psychiatry, 34(4), 323–331. https://doi.org/10.1016/j.genhosppsych.2012.02.007
47. Mullinax LRA, Grunwald L, Banaag A, et al. A longitudinal study of prevalence ratios for musculoskeletal back injury among U.S. Navy and Marine Corps personnel, 2009-2015. Mil Med 2023;188(5–6): e1094–e1101.
48. Plesh O, Adams SH, Gansky SA. Racial/Ethnic and gender prevalences in reported common pains in a national sample. Journal of orofacial pain 2011 25(1):25-31.
49. Ezenwa MO, Fleming MF. Racial disparities in pain management in primary care. J Health Dispar Res Pract 2012;5(3):12–26.
50. Hoffman KM, Trawalter S, Axt JR, et al. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. Proc Natl Acad Sci U S A 2016;113(16):4296–4301.
51. Casazza BA. Diagnosis and treatment of acute low back pain. Am Fam Physician 2012;85(4):343–350.

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 154–163
DOI 10.1002/acr.25651

American College of Rheumatology
Empowering Rheumatology Professionals

Assessment of Pain Types in Recently Diagnosed Patients With Inflammatory Arthritis

Zoe Rutter-Locher,[1,2] Sam Norton,[2] Joseph L. Taylor,[2] Bina Menon,[1] Tom Esterine,[2] Ruth Williams,[2] Leonie S. Taams,[2] Kirsty Bannister,[3] and Bruce W. Kirkham[1]

Objective. Up to 40% of patients with inflammatory arthritis (IA) experience persistent pain, traditionally thought to be associated with a shift from peripherally to centrally mediated pain during the disease course in some patients. We assessed sensory profiles of recently diagnosed individuals with IA, hypothesizing that pain reported at this early stage of diagnosis is driven predominantly by peripheral joint inflammation.

Methods. Recently diagnosed patients with IA with pain numerical rating scale scores of ≥3 were recruited. We collected data on the following: arthritis activity (Disease Activity Score in 28 joints [DAS28], musculoskeletal ultrasonography), quality of life (Musculoskeletal Health Questionnaire [MSK-HQ], EuroQol 5-domain), mental health status (Patient Health Questionnaire Anxiety–Depression Scale [PHQ-ADS]), and pain characteristics (fibromyalgia criteria, painDETECT, static and dynamic quantitative sensory testing [QST]).

Results. Sixty-one participants (57% female, 62% with rheumatoid arthritis) were enrolled (mean ± SD age 49.8 ± 15 years; mean ± SD time since diagnosis 1.2 ± 2.3 months). Ninety-seven percent had peripheral joint inflammation, with a mean ± SD DAS28 score of 3.8 ± 1. However, 21% met the fibromyalgia criteria, 25% had a painDETECT score of ≥19, and 20% had a tender joint count minus swollen joint count of ≥7, which significantly correlated with DAS28, MSK-HQ, and PHQ-ADS scores. QST revealed lowered pressure pain thresholds at nonarticular sites in a subset of participants and facilitated temporal pain summation and deficient pain modulation in 18% and 61% of patients, respectively.

Conclusion. This study provides evidence of centrally mediated pain at the time of diagnosis, challenging the notion that, even at the early stage of disease, pain is driven only by peripheral mechanisms.

INTRODUCTION

Despite major improvements in therapies and treatment strategies, up to 40% of patients diagnosed with inflammatory arthritis (IA), including rheumatoid arthritis (RA), psoriatic arthritis, and ankylosing spondylitis, continue to experience persistent pain.[1–4] Pain intensity initially improves during the six-month period post diagnosis, likely attributable to the anti-inflammatory properties of disease-modifying antirheumatic drugs. However, after that period, joint tenderness and pain persist, which might be unresolved inflammation, or as often found in the absence of synovitis, it might be pain sensitization[1] or a combination of both.[5] A recent meta-analysis by our group found that, according to patient-reported outcome measures (PROMs), pain sensitization was present in 20% to 30% of individuals with established IA, with the largest group being those with psoriatic arthritis.[6] What underlies this pain sensitization remains equivocal, with postulated mechanisms encompassing dysfunction in both peripheral and central pain-processing pathways, including spinal and supraspinal circuits.[7–10] Improving our understanding of how persistent pain is initiated, develops, and is maintained in the patient population with IA is crucial because it could help guide appropriate therapeutic strategies.

The views expressed in this publication are those of the authors and do not necessarily represent those of the NIHR, the NHS, or the UK Department of Health and Social Care.

Dr Rutter-Locher's work is supported by the NIHR for this research project (Doctoral Fellowship, NIHR301674).

[1]Guy's and St Thomas' NHS Trust, London, United Kingdom; [2]King's College University, London, United Kingdom; [3]Imperial College London, London, United Kingdom.

Drs Bannister and Kirkham were co-senior authors and contributed equally to this work. Mr Esterine and Dr Williams contributed as patient research partners.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25651).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25651.

Address correspondence via email to Zoe Rutter-Locher, MSc, at Zrutter-locher@nhs.net.

Submitted for publication March 31, 2025; accepted in revised form September 2, 2025.

SIGNIFICANCE & INNOVATIONS

- This study is the first to record comprehensive sensory profiles incorporating clinical examination, ultrasonography, questionnaire, and quantitative sensory testing (QST) data in the early stage of inflammatory arthritis diagnosis.
- Of those tested, 97% had peripheral joint inflammation confirmed by clinical assessment and ultrasonography. In addition, clinical, patient-reported, and QST outcome measures revealed the likely presence of centrally mediated pain in a subset of participants.
- These findings challenge the longstanding assumption that only peripheral mechanisms driven by joint inflammation underpin pain at diagnosis, highlighting a role for centrally mediated pain even in the early stage of the disease.

Pain in IA can have many different causes, ranging from persistent joint inflammation to dysfunctional neural processes, which require the use of multiple assessment tools to provide a more complete understanding. Musculoskeletal ultrasonography (MSKUS) provides an objective assessment of peripheral joint inflammation in addition to clinical examination, whereas quantitative sensory testing (QST) allows activity in peripheral nerves to be inferred (static QST) alongside the functionality of central pain-processing pathways (dynamic QST). When combined with validated questionnaire data, which provide insights into the patient pain experience, clinical examination, MSKUS, and QST outcomes can be used to build an individual's sensory profile. Cumulatively, these profiles can help deduce likely pain-causing mechanisms.[11,12]

Studies using questionnaire and QST methods, mainly cross-sectional design in patients with longstanding IA,[13,14] have suggested the presence of centrally mediated pain in a subset of patients. This research has led to the current consensus that, in some patients, during their disease course, pain primarily caused by peripheral joint inflammation at diagnosis transforms to include pain that is underpinned by dysfunction in central nervous system processes.[7] The temporal nature of this shift during the disease course remains unclear because no studies have explored detailed pain characteristics at diagnosis, and there are limited longitudinal studies. In keeping with the current understanding that centrally mediated pain develops over the course of the disease, we hypothesized that peripherally mediated inflammatory pain would predominate during this initial disease phase, with little evidence of dysfunctional central pain processing. Here we report the initial sensory profiles of individuals with IA promptly after diagnosis, established with an array of techniques encompassing both MSKUS and QST alongside validated questionnaires. To our knowledge, this is the first study to employ such a detailed multimodal approach to sensory profiling in newly diagnosed patients with IA.

PATIENTS AND METHODS

Study population. This study presents baseline data obtained from an initial cohort of 61 patients enrolled in the Pain Phenotypes and Their Underlying Mechanisms in Inflammatory Arthritis Study (PUMIA). PUMIA is a single-site, prospective, longitudinal, observational study aiming to understand pain mechanisms in IA. Recruitment of participants took place at the rheumatology outpatient department of Guy's Hospital, London, between February 2022 and November 2022. Inclusion criteria for the study encompassed individuals who reported a total numerical rating scale (NRS) pain score of 3 or higher on a scale of 0 to 10, received a clinician diagnosis of peripheral inflammatory joint disease, and experienced symptom onset within the past 12 months. Patients were enrolled in the study as soon as feasible following their diagnosis.

Patients were excluded if they were under 18 years of age, were unable or unwilling to provide informed written consent, were unable to adhere to study protocols, were pregnant or breastfeeding, were receiving non-IA–related immunosuppressant therapy, were undergoing current or recent (within the last 90 days) treatment with investigational agents, or had a history of severe peripheral vascular disease or peripheral neuropathy. Use of opioids, gabapentin, or pregabalin was stopped 24 hours before assessment. All participants gave fully informed written consent, and the study protocol received approval from Bromley Research Ethics Committee and the Health Research Authority (research ethics committee number 21/LO/0712).

Clinical assessment and ultrasonography. Routine demographic data (including age, sex, ethnicity—self-reported from a predefined list—body mass index, smoking status, employment status, and comorbidities), disease data (including diagnosis, disease duration, and serology), and medication history were obtained from each participant. The term "current steroids" refers to oral prednisolone taken within the past two weeks or intramuscular methylprednisolone acetate administered within the last three months.

Patients were assessed for disease activity (C-reactive protein [CRP], 68 tender joint count [TJC], 66 swollen joint count [SJC], Disease Activity Score in 28 joints [DAS28], TJC – SJC using 28 joints) and completed questionnaires for quality of life (QoL; Musculoskeletal Health Questionnaire, EuroQol 5-domain[15]), impact of disease (Rheumatoid Arthritis Impact of Disease[16]), somatic symptoms (Patient Health Questionnaire 15 [PHQ-15][17]), and mental health (PHQ-9 for depression,[18] Generalized Anxiety Disorder 7 [GAD-7][19]). Fibromyalgia status was determined according to the 2016 revision of the American College of Rheumatology 2010 modified fibromyalgia diagnostic

criteria[20] (Widespread Pain Index [WPI] and Symptom Severity Scale [SSS] score). Fibromyalgia severity was assessed as the sum of the WPI and SSS scores. Pain assessment was conducted using (1) the NRS for, pain in which individuals are asked to score, between 0 and 10, both total body pain and most painful joint pain in the last 24 hours (the "target joint"), and (2) the painDETECT questionnaire, which identifies neuropathic-like pain.[21] The painDETECT has been used in previous studies as a proxy for pain caused by central mechanisms due to the similarity in symptom profiles.[22–24]

A 14-joint MSKUS assessment was used to obtain an objective evaluation of joint inflammation, particularly subclinical synovitis. Bilateral radiocarpal joints, intercarpal joints, ulnocarpal joints, metacarpophalangeal joints 2 and 3, proximal interphalangeal joints 2 and 3, and metatarsophalangeal joints 2 and 3 were graded semiquantitatively for gray scale (GS) and power Doppler (PD) according to the EULAR–Outcome Measures in Rheumatology criteria[25] by two individual raters, blinded to patient demographics or sensory profile.

QST. *Static QST.* The QST protocols established by the German Research Network on Neuropathic Pain (DFNS) were applied to assess cutaneous sensory detection using standardized von Frey hairs (mechanical detection threshold [MDT]) and pain thresholds using a standardized pinprick set (mechanical pain threshold [MPT]) on the target most painful joint and control site (contralateral volar forearm).[26] These tests are performed using standardized instructions in which normative data, grouped by age, sex, and site, are available to allow comparisons with healthy controls.[27] The established DFNS protocol was also used to assess the pressure pain threshold (PPT) at the joint line of the target most painful joint, the dorsal aspect of bilateral wrists, and a nonarticular control site (bilateral trapezius muscles) using a Wagner Force 25 FDX algometer (Wagner Instruments).[26] PPT is the point at which a pressure applied first becomes painful, and lower pain thresholds reflect higher pain sensitivity. Lower PPT at nonarticular sites suggest widespread pain likely underpinned by central mechanisms, whereas lower PPT at joint sites only suggests local joint pain sensitivity due to joint inflammation and peripherally mediated mechanisms.

Dynamic QST. Dynamic QST comprises temporal summation of pain (TSP) and conditioned pain modulation (CPM) paradigms. The TSP paradigm provides a proxy measure of spinal neuronal activity. In brief, delivery of an equal-intensity noxious stimulus above a critical rate leads to an increase in the individual's perception of that stimulus. We used two (TSP) paradigms. The first followed a DFNS protocol using pinprick stimulators,[26] and the second followed a standard protocol using cuff pressure algometry (detailed protocol in the Supplementary Material). Spinal neuronal activity can be modulated through descending modulatory circuits, whose function can be assessed upon application of a CPM paradigm.[28,29] In brief, a measure of painfulness for a test stimulus is taken, in the absence or presence of a second painful "conditioning" stimulus. The difference in the participant's ratings of the test stimuli (before and during conditioning) is the measure of a "CPM effect," in which a "CPM responder" refers to an individual who experiences a significant reduction in pain rating of the test stimulus upon conditioning whereas a "CPM nonresponder" does not. For the CPM paradigm, we followed a computerized cuff algometry protocol.[30]

Statistical analysis. All data are presented as mean ± SD. Paired or independent *t*-tests were employed, as appropriate, to identify significant differences in means. To normalize the data for MDT, MPT, PPT, and TSP, logarithmic transformations were applied,[27,31] resulting in values with a mean of 0 and an SD of 1. Absolute reference data (matched for age, sex, and site—dorsum of the hand for MDT, MPT, and TSP and thenar eminence for PPT) were used to normalize test results of the individual participant by calculating the Z transformation: (value participant – mean controls)/SD controls.[32] A Z score of ±1.96 corresponded to statistical significance at the 0.05% level.

In addition to descriptive analysis, to provide a more formal test of the relative contribution of inflammation and other candidate variables to pain, we conducted a regression model with pain NRS as the dependent variable. MSKUS total GS and PD; our preferred measure of peripheral inflammation) was entered in the first block, followed by demographic factors in the second block to estimate the adjusted association, followed by other plausible contributors to pain, including fatigue, mood, and centrally mediated pain markers, in a third block. This blockwise approach allowed us to estimate the variance in pain explained by inflammation alone and the incremental contribution of other factors.

Pearson correlations were used to investigate the relationships between markers of central pain mechanisms, peripheral inflammation, and QST parameters and clinical outcomes, such as pain levels, disease activity, and mental health. All assumptions for parametric tests were checked before analysis. For variables showing deviation from normality, key analyses were repeated using nonparametric equivalents (eg, Spearman's correlation) as a sensitivity check. Given the exploratory nature of this analysis, effect sizes were primarily evaluated. For all analyses, *P* values below 0.05 were considered statistically significant.

Finally, an exploratory factor analysis was conducted to identify underlying latent variables (factors) accounting for the common variance across observed variables (detailed protocol in the Supplementary Material). All analyses were performed using Stata version 17.0 (StataCorp LLC). All data are available upon request to the corresponding author.

RESULTS

Demographics and clinical assessment. *Demographics.* We recruited 61 patients composed of a broad range of

ethnicities predominantly with a diagnosis of RA (62%), with the remaining 38% having a mixture of early IA, psoriatic arthritis, and spondylarthritis. The mean age of participants was 49.8 years (range 19–86 years), and most participants were female (57%). Mean ± SD symptom duration was 6.3 ± 3.1 months, and mean ± SD time since diagnosis was 1.2 ± 2.3 months (Table 1).

Disease activity and therapy. The mean ± SD DAS28-CRP was 3.8 ± 1. The mean ± SD total and target joint pain on an NRS (0–10) was 5.5 ± 2.1 and 5.1 ± 2.5, respectively (Table 1).

Mental health and QoL. The prevalence of participants with patient-reported outcome indications of anxiety, depression, and somatization in the patient population was high (Table 1 and Figure 1). QoL was in keeping with established RA populations.[33]

Ultrasonographic assessment. There was evidence of active synovitis either clinically or on ultrasonography in 59 patients (97%), in keeping with a recent IA diagnosis. Of the 59, 37 (61%) had GS positivity, and 22 (36%) had PD positivity at the target joint (Table 2). Of the 24 patients (39%) who had neither GS or PD positivity at the target joint, 17 had GS or PD positivity at other joints, and 7 did not, although 5 of those had clinically swollen joints not examined by ultrasonography. Of note, 24 of the 39 patients with PD 0 at the target joint had recently received steroid therapy, which may have suppressed the PD signal.

Clinical outcome measures and PROMs. The assessment of WPI and TJC – SJC provides an indication of widespread pain,[34] which is often associated with centrally mediated pain. Of the total 61 participants, 20% to 30% fulfilled criteria for widespread pain, fibromyalgia, and neuropathic-like pain on the painDETECT questionnaire (Table 1), and these PROMs were correlated (Supplementary Table S1).

QST. In this section, we provide a summary of the QST findings. Further details for interested readers are presented in the Supplementary Material.

Static QST: cutaneous sensation and pain thresholds. Cutaneous sensation (MDT) and pain thresholds (MPT) showed no deviation from normative values at the target joint or control site (Figure 2A, Z values < 1.96). There was a significant increase in sensitivity in MPT at the control site compared to the target joint (P = 0.006), which may reflect a normal difference in sensation between the joint and forearm sites or hypoesthesia due to inflammation at the joint (Supplementary Table S2). Comparison with healthy controls would be required to determine this.

Static QST: PPT at target joint and trapezius. A low PPT reflects pain sensitization, and a higher Z score is indicative of gain of function (sensitivity), with a Z score of ≥1.96 representing significant sensitivity. As a group, participants showed significantly increased sensitivity compared to normative data at all sites except the right trapezius (mean PPT Z score at target joint 3.8 [SD 3.0], left wrist 2.9 [SD 3.1], right wrist 2.9 [SD 2.5], left trapezius 2.1 [SD 2.5], right trapezius 1.7 [SD 2.4]; Supplementary Table S3). Sensitivity was significantly greater at the target joint than at the left (P = 0.007) and right trapezius (P = 0.000). Sixteen (26%) had significant sensitivity (Z score ≥ 1.96) at joint sites only, suggesting peripheral sensitization. Thirty-four (56%) had significant sensitivity at both joint and trapezius sites, indicating possible widespread centrally mediated pain. No patients had trapezius sensitivity without joint or wrist sensitivity. Eleven (18%) showed no significant sensitivity at any site.

Dynamic QST: TSP. Using the cuff algometer, 9 of 51 (18%) fulfilled criteria for facilitated TSP with a ratio of >2.48 (Figure 3A and Supplementary Figure S1). In contrast, there was no difference in TSP, using pinpricks compared to normative data at either the target joint (test) or the control site (Figure 3B).

Dynamic QST: CPM. In total, 35 of 57 participants (61%) were CPM nonresponders (ie, those not reporting a decrease in pain perception upon the application of a painful test stimulus concurrent to a conditioning stimulus) based on pain detection thresholds, and 44 of 57 (77%) were CPM nonresponders based on pain tolerance thresholds (Figure 3C and D). Although pain detection thresholds did not vary across CPM nonresponder and responder groups, CPM nonresponders reported lower pain tolerance thresholds (Supplementary Table S4), highlighting that although both groups had similar initial pain sensitivity, nonresponders had a reduced ability to endure pain.

Multivariable linear regression model. In blockwise linear regression, MSKUS total GS and PD alone (block 1), our measure of peripheral inflammation, explained little of the variance in pain scores (R^2 = 0.0006, P = 0.86). Adding age and sex (block 2) modestly increased the explained variance (R^2 = 0.0619, P = 0.36). In the final model (block 3), which included additional variables plausibly linked to pain, fatigue, mood (PHQ Anxiety–Depression Scale), fibromyalgia severity (WPI and SSS), painDETECT score, and trapezius PPT, the explained variance increased substantially (R^2 = 0.3538, P = 0.009). Contrary to our hypothesis that inflammation would be a key driver of pain, MSKUS total GS and PD contributed little to the variance in pain scores, suggesting that noninflammatory factors may contribute more to pain severity in this cohort.

Correlations and factor analysis of clinical outcome measures and PROMs. Clinical and patient-reported outcomes that suggest centrally mediated pain and lower nonarticular PPT correlated with higher disease activity (DAS28), increased fatigue, decreased QoL, greater impact of disease, and worse mental health measures (Table 3). In contrast, clinical measures of peripheral inflammation (total GS and PD positivity on ultrasonography, SJC) did not (Table 3). There was no correlation between clinical markers of central pain and dynamic QST (Supplementary Table S5).

Table 1. Demographic and pain data*

Variable	Value
Demographic (N = 61)	
Age, mean ± SD (range), years	49.8 ± 15.4 (19–86)
Female sex, n (%)	35 (57)
Ethnicity, n (%)	
White British	33 (54)
Black or Black British	11 (18)
Asian or Asian British	7 (12)
Other	10 (16)
Smoker, current or previous, n (%)	23 (38)
Unemployed, n (%)	9 (15)
BMI, mean (SD)	26.8 (5.7)
Diagnosis	
EIA, n (%)	16 (26)
RA, n (%)	38 (62)
PsA, n (%)	6 (10)
SpA, n (%)	1 (2)
Seropositive: RF or CCP, n (%)	37 (61)
Time since diagnosis, mean (SD), months	1.2 (2.3)
Symptom duration, mean (SD), months	6.3 (3.1)
DAS28-CRP, mean (SD)	3.8 (1.0)
TJC-28, mean (SD)	6.5 (5.2)
SJC-28, mean (SD)	2.7 (2.5)
Steroid and DMARDs, n (%)	
Steroids in some form at assessment	37 (61)
IM depomedrone in last 3 months	25 (41)
PO prednisolone in last 2 weeks	13 (21)
IV methylprednisolone	2 (3)
cDMARDs	36 (59)
Monotherapy	33 (54)
Combination therapy	3 (5)
Methotrexate	32 (52)
Quality of life, mean (SD)	
MSK-HQ	29 (11)
EQ-5D	0.58 (0.28)
EQ-5D total health VAS	59 (22)
RAID	4.7 (2.6)
Mental health	
GAD-7, mean (SD)	7.4 (6.0)
Anxiety total, score ≥5, n (%)	38 (62)
Mild anxiety, score 5–9, n (%)	21 (34)
Moderate anxiety, score 10–14, n (%)	7 (11)
Severe anxiety, score ≥15, n (%)	10 (16)
PHQ-9, mean (SD)	8.1 (6.4)
Depression total, score ≥5, n (%)	40 (66)
Mild depression, score 5–9, n (%)	17 (28)
Moderate depression, score 10–14, n (%)	13 (21)
Moderately severe depression, score 15–19, n (%)	7 (11)
Severe depression, score ≥20, n (%)	3 (5)
PHQ-15, mean (SD)	9.0 (5.5)
Somatic symptoms total, score ≥5, n (%)	49 (80)
Mild somatic symptoms, score 5–9, n (%)	25 (40)
Moderate somatic symptoms, score 10–14, n (%)	15 (24)
Severe somatic symptoms, score ≥15, n (%)	9 (15)
Pain, mean (SD)	
Total body pain	5.5 (2.1)
Target joint pain	5.1 (2.5)
Widespread pain	
WPI score, mean (SD)	5.6 (2.9)
WPI, score of ≥7, n (%)	19 (31)
Fibromyalgia, n (%)	13 (21)
TJC − SJC, score of >7, n (%)	12 (20)
painDETECT	
painDETECT, mean (SD)	12.7 (8.0)
Likely neuropathic-like pain, score of ≥19, n (%)	15 (25)
Possible neuropathic-like pain, score between 13 and 18, n (%)	9 (15)
Unlikely neuropathic-like pain, score ≤12, n (%)	37 (61)

Table 1. (*Cont'd*)

* Table showing patient demographics, diagnosis, and medication alongside results of questionnaire data collected for quality of life, mental health, fibromyalgia criteria, and painDETECT (N = 61). BMI, body mass index; CCP, cyclic citrullinated peptide; cDMARD, conventional disease-modifying antirheumatic drug; DAS28-CRP, Disease Activity Score in 28 joints using the C-reactive protein level; DMARD, disease-modifying antirheumatic drug; EIA, early inflammatory arthritis; EQ-5D, EuroQol 5-domain; GAD-7, Generalized Anxiety Disorder 7; IM, intramuscular; IV, intravenous; MSK-HQ, Musculoskeletal Health Questionnaire; PHQ, Patient Health Questionnaire; PO, oral; PsA, psoriatic arthritis; RA, rheumatoid arthritis; RAID, Rheumatoid Arthritis Impact of Disease; RF, rheumatoid factor; SJC-28, swollen joint count for 28 joints; SpA, spondylarthritis; TJC-28, tender joint count for 28 joints; TJC − SJC, tender joint count minus swollen joint count; VAS, visual analog scale; WPI, Widespread Pain Index.

Based on factor analyses of the 14 variables, we identified two underlying latent factors. CRP, TSP, and CPM had Kaiser–Meyer–Olkin (KMO) statistics below 0.5 and were not included in the model. The overall KMO statistic was 0.76, indicating the correlation matrix of the remaining variables was appropriate for factor analysis.[35] Factor 1 included self-reported variables of pain (total pain, patient global assessment, TJC, painDETECT, fibromyalgia criteria, and trapezius PPT) and fatigue and mental health (PHQ-9 and GAD-7). Factor 2 included clinical and ultrasonographic measures of synovitis (SJC, GS and PD on MSKUS, and physician global assessment; Supplementary Table S6). The correlation between the two factors was –0.13.

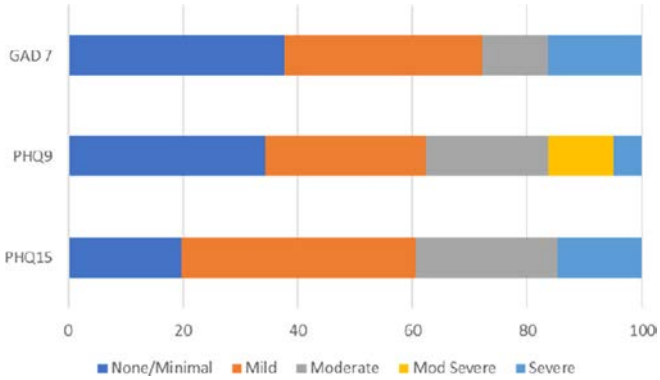


Figure 1. Prevalence of anxiety (GAD 7), depression (PHQ9), and somatic symptoms (PHQ15) in the study population (N = 61). GAD 7, Generalized Anxiety Disorder 7; PHQ, Patient Health Questionnaire.

Table 2. GS and PD activity on musculoskeletal ultrasonography of the target most painful joint*

GS score	n (%)	PD score	n (%)
GS 0	24 (39)	PD 0	39 (64)
GS 1	16 (26)	PD 1	5 (8)
GS 2	18 (30)	PD 2	11 (18)
GS 3	3 (5)	PD 3	6 (10)

* n (%) values of patients with each score at the target joint. GS, gray scale; PD, power Doppler.

DISCUSSION

This study provides novel insights into the mechanisms underlying pain reported by patients with early IA, as we compiled comprehensive sensory profiles of individuals shortly after diagnosis. Consistent with newly diagnosed IA, 97% of patients had active synovitis, confirming that peripheral inflammation is highly prevalent at this stage and likely contributes to pain. However, we also observed indicators of centrally mediated pain in approximately 20% to 30% of patients who reported widespread pain, fulfilled diagnostic criteria for fibromyalgia, and/or demonstrated elevated painDETECT scores. QST findings further supported the involvement of centrally mediated pain mechanisms. Reduced PPTs at nonarticular sites in some patients suggested widespread hyperalgesia, whereas heightened responses to TSP and impaired CPM indicated dysfunction in spinal and supraspinal processes. The regression analysis showed that inflammation alone explained very little variance in total pain scores, and it was centrally mediated pain factors that correlated more strongly with clinical outcomes such as QoL and impact of disease. This suggests that although inflammation may trigger nociceptive input, other factors, including central mechanisms driven by, for example, psychosocial factors, play a substantial role in shaping individual pain experiences even early in the disease course. Overall, these data challenge the longstanding assumption that only peripheral mechanisms driven by joint inflammation underpin pain at diagnosis. Instead, they suggest that some patients with early IA have both inflammatory and centrally mediated pain, similar to patients with established IA.[5,6,14]

Our cohort's clinical disease activity measures and average pain score of 5.5 of 10 aligns with findings reported in other early IA cohorts,[36] as do the comparably high rates of anxiety and/or depression.[37] Although pain mechanisms in early IA remain relatively unexplored, our findings revealed that 21% of patients met the criteria for fibromyalgia. Some of these cases may represent individuals with underlying primary fibromyalgia. However, this figure is much higher than the estimated prevalence of 1.78% in the general population.[38] In addition, the mean symptom duration of six months would allow sufficient time for, for example, centrally mediated pain to develop in response to peripheral nociceptive input. It also exceeds the 5.2% prevalence reported in one Canadian study.[39] This may reflect methodologic differences, with the Canadian study relying on clinical judgment rather than strict fibromyalgia diagnostic criteria or the Canadian cohort having a more inflammatory profile than ours or other early IA cohorts.[40,41]

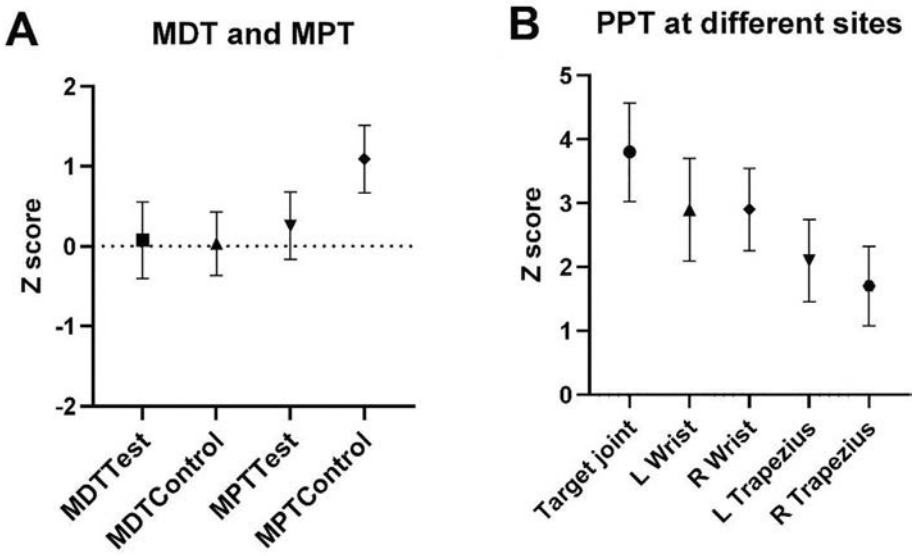


Figure 2. Graphs showing results from static QST. (A) Z score for the study population MDT and MPT at test (target joint) and control sites (contralateral forearm) (n = 61). The QST Z score chart illustrates the trend of sensory change in which positive Z scores suggest increased sensory capability (ie, hypersensitivity), whereas negative scores point to diminished sensory ability (ie, hyposensitivity). There is no significant difference in the MDT or MPT compared to normative data, reflected by Z scores <1.96. (B) Z score for PPT at each site (n = 60). As a group, participants showed significantly increased sensitivity at all sites, except the right trapezius (mean PPT at target joint 3.8 [SD 3.0], left wrist 2.9 [SD 3.1], right wrist 2.9 [SD 2.5], left trapezius 2.1 [SD 2.5], right trapezius 1.7 [SD 2.4]). Sensitivity was significantly greater at the target joint than at the left (P = 0.007) and right trapezius (P = 0.000). L, left; MDT, mechanical detection threshold; MPT, mechanical pain threshold; PPT, pressure pain threshold; QST, quantitative sensory testing; R, right.

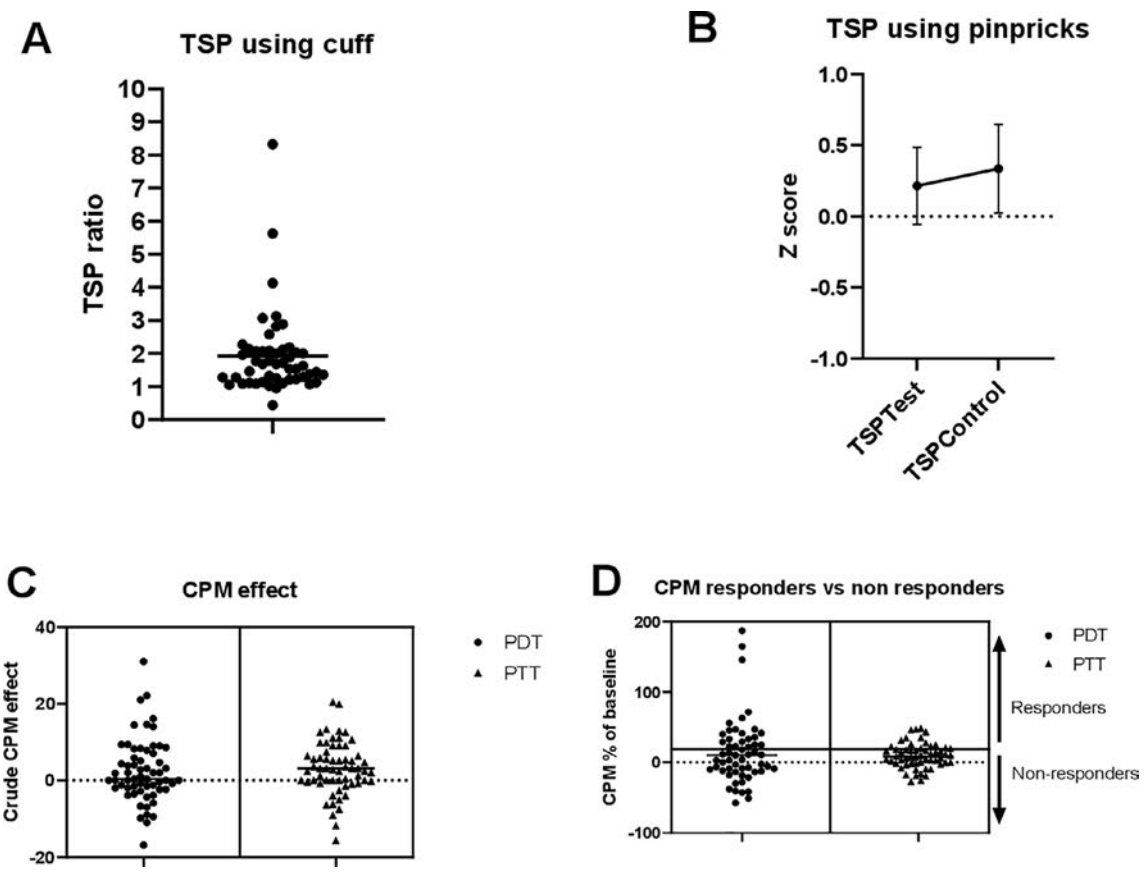


Figure 3. Graphs showing results from dynamic QST. (A) Individual values for the TSP using cuff (n = 51). Nine of fifty-one (18%) fulfilled the criteria for facilitated TSP, with a ratio of >2.48. (B) Z score for the study population TSP using pinpricks (n = 51). There was no significant difference in TSP using pinpricks compared to normative data, reflected by Z scores <1.96. (C) Graph showing individual values of crude CPM effect ($PDT/PTT_{(CPM)} - PDT/PTT_{(Dom)}$) for PDT and PTT (n = 57). (D) Graph showing individual values of CPM effect as a percentage of baseline for PDT and PTT. Participants were stratified as responders (CPM response >20% of the baseline cPPT) or nonresponders (CPM response ≤20% of baseline cPPT) (n = 57). According to PDT and PTT, 35 of 57 (61%) and 44 of 57 (77%), respectively, were nonresponders. CPM, conditioned pain modulation; cPPT, cuff pressure pain threshold; Dom, Dominant leg; PDT, pain detection threshold; PTT, pain tolerance threshold; QST, quantitative sensory testing; TSP, temporal summation of pain.

In addition, symptoms of very active early IA, such as widespread pain, fatigue, and unrefreshing sleep, overlaps with fibromyalgia criteria, potentially resulting in a "false" fibromyalgia diagnosis. However, the lack of correlation between WPI/fibromyalgia severity and peripheral inflammation markers suggests that the observed widespread pain may be driven by mechanisms other than joint inflammation.

We observed reduced PPT values at the affected joint in most patients, indicating local pain sensitization, likely due to joint inflammation, in this early arthritis cohort. In addition, we found lowered PPT values at nonarticular sites, indicative of widespread pain, in some patients. These findings are in keeping with those from a recent meta-analysis of patients with established IA[14] but are the first to identify such reductions in PPT at nonarticular sites specifically in the context of early IA, highlighting the presence of widespread pain sensitivity in some patients early in the disease course. Factor analysis revealed close links between reduced PPT at nonarticular sites and self-reported pain, fatigue, and poor mental health, which are often associated with centrally mediated pain. The lack of interaction with dynamic QST (TSP or CPM) measures suggests that reduced PPT at nonarticular sites may be the most robust QST indication of centrally mediated pain in this patient population. Although the analysis was used primarily as a data reduction technique rather than to identify a replicable latent structure, it was conducted on a relatively small sample. This limits the stability and generalizability of the resulting factor structure.

Eighteen percent of participants demonstrated facilitated TSP, in which a TSP paradigm with the cuff algometer was used as a repetitive stimulus, indicating spinal mechanisms contributing to heightened pain. However, when pinpricks were used as the repetitive stimulus, facilitated TSP was not observed, contrasting with findings from a study that also used pinpricks in patients with established IA.[42] This difference in findings could be attributed to characteristics specific to our study population—ours was a recently diagnosed patient cohort with

Table 3. Correlation between markers of centrally mediated pain/peripheral inflammation and clinical parameters*

	DAS28	TJC-28	SJC-28	Time since diagnosis	Symptom duration	Fatigue	MSK-HQ	RAID	GAD-7	PHQ-9	PHQ-15
Markers of centrally mediated pain											
Total pain	0.513[a,e]	0.394[b,e]	0.006[d]	−0.103[c]	0.116[c]	0.460[b,e]	−0.475[b,e]	0.58[a,e]	0.301[b,e]	0.307[b,e]	0.386[b,e]
Fibromyalgia severity (WPI + SSS)	0.532[a,e]	0.524[a,e]	−0.093[d]	−0.153[c]	0.039[d]	0.647[a,e]	−0.653[a,e]	0.684[a,e]	0.636[a,e]	0.685[a,e]	0.674[a,e]
painDETECT	0.514[a,e]	0.454[b,e]	−0.132[c]	−0.175[c]	−0.073[d]	0.501[a,e]	−0.5[a,e]	0.596[a,e]	0.501[a,e]	0.493[b,e]	0.681[a,e]
PPT at the trapezius (average)	−0.293[c]	−0.267[c]	0.082[d]	0.043[d]	0.098[d]	−0.301[b]	0.333[b,e]	−0.432[b]	−0.188[c]	−0.295[c]	−0.286[c]
Markers of peripheral inflammation											
SJC	0.337[b,e]	0.207[c]	–	0.038[d]	0.077[d]	0.022[d]	−0.112[c]	0.032[d]	−0.243[c]	−0.191[c]	−0.125[c]
Total GS and PD positivity	−0.106[c]	−0.007[d]	0.546[a,e]	−0.019[d]	−0.001[d]	−0.021[d]	−0.036[d]	−0.039[d]	−0.210[c]	−0.222[c]	−0.199[c]

* Correlation analysis for measures of centrally mediated pain and peripheral inflammation with clinical parameters. Each cell displays the correlation coefficient between a pain or inflammation marker (rows) and a clinical variable (columns). Numerical correlation coefficients are annotated within each cell. Higher total pain, fibromyalgia severity, and painDETECT scores and lowered PPT at the bilateral trapezius (taken as an average of both sides) correlated with higher disease activity (DAS28) and TJC but not SJC. They also correlated with increased fatigue, decreased QoL (MSK-HQ), and greater impact of disease (RAID). Additionally, these measures showed correlations with mental health measures, including higher levels of anxiety (GAD-7), depression (PHQ-9), and somatic symptoms (PHQ-15) (with the exception of PPT at the trapezius and anxiety). Notably, these measures did not demonstrate correlations with either time since diagnosis or symptom duration. Total GS and PD positivity in the 14-joint ultrasonography assessment represents the combined scores of GS and PD assessments. The SJC and total GS and PD positivity serve as indicators of peripheral inflammation, which contributes to peripheral pain. SJC correlated with disease activity (DAS28), and total GS and PD positivity correlated with SJC. However, neither the SJC nor total GS and PD positivity correlated with TJC, fibromyalgia severity, QoL by MSK-HQ or RAID, or mental health measures (GAD-7, PHQ-9, or PHQ-15). DAS28, Disease Activity Score in 28 joints; GAD-7, Generalized Anxiety Disorder 7; GS, gray scale; MSK-HQ, Musculoskeletal Health Questionnaire; PD, power Doppler; PHQ, Patient Health Questionnaire; PPT, pressure pain threshold; QoL, quality of life; RAID, Rheumatoid Arthritis Impact of Disease; SJC, swollen joint count; SSS, Symptom Severity Scale; TJC, tender joint count; WPI, Widespread Pain Index.

[a] Strong correlations ($|r| \geq 0.5$).

[b] Moderate correlations ($|r| = 0.3–0.49$).

[c] Weak correlations ($|r| = 0.1–0.29$).

[d] Negligible correlations ($|r| < 0.1$).

[e] Statistical significance at the $P < 0.05$ level.

IA—or, alternatively, due to methodologic protocols. For example, the initial pinprick force may not have been sufficiently noxious, or the initial pinprick rating may have been too high, creating a ceiling effect on subsequent pain ratings. When applying a CPM paradigm, 61% and 77% of the participants exhibited a "nonresponder status," according to pain detection threshold and pain tolerance threshold, respectively, indicating dysfunction in a naturally occurring pain inhibitory pathway. Although data are lacking to compare the proportion of nonresponders with healthy controls,[29] our cohort's mean CPM effect for pain detection threshold was 2.4, below that found in healthy populations, in which the mean CPM effect for pain detection threshold ranges[30] from 5.3 to 11.5.

This study has several strengths. The in-depth sensory profiling allowed us to create a comprehensive profile, combining measures of pain sensitivity and centrally mediated pain-processing mechanisms and ultrasonography as an objective measure of joint inflammation. All assessments were performed by a single researcher, ensuring consistency and minimizing interrater variability. We attempted early assessment of patients soon after their diagnosis, which was usually achieved, enabling insights into what is driving pain in the very early stages of IA soon after diagnosis.

The study also has limitations. One key limitation is that there is currently no gold standard comprehensive method for assessing centrally mediated pain. As a result, instruments such as questionnaires and QST, which have been validated individually, are combined. However, these instruments are frequently not standardized or validated in comparison to each other. For example, although used in the literature as a proxy measure of centrally mediated pain, there is currently limited data linking "likely neuropathic pain" identified by painDETECT to QST findings in IA.[43] Furthermore, as PPT normative data for the wrist and trapezius were not available, data from the thenar eminence (a potentially more sensitive site) were used for comparison. Because PPT varies by anatomic site, this may have led to systematically inflated Z scores at the wrist and trapezius and could partly explain the high proportion of participants showing sensitivity. Alternatively, it is possible that PPT is a more sensitive measure of pain processing than PROMs.

Another important consideration is that the enrollment criteria specified pain levels of ≥3. This threshold was chosen to align with the study's aim of exploring pain mechanisms in early IA. However, these criteria may limit the generalizability of results to those people in at least moderate pain, and it is likely that, in the absence of this threshold, the proportion of participants meeting criteria for centrally mediated pain would be lower. Additionally, given the relatively short time since diagnosis, it is possible that some participants would not go on to develop chronic IA, which would limit the generalizability of the results. At the time of study assessment, most patients had already received therapy, and although we captured patients within a short time post diagnosis, most individuals had pain in the period before diagnosis, and so in some individuals, central nervous system dysfunction may have predated joint inflammation. It has been postulated that a preloading psychological vulnerability to pain may contribute to the development of persistent noninflammatory pain. An alternative explanation is that this could also be influenced by subclinical joint pathology. We will follow this group with longitudinal investigations aimed at identifying evolution of pain mechanisms over time and possible signals for those who might develop persistent noninflammatory pain.

These findings emphasize that pain in early IA is complex and, in some people, includes centrally mediated pain mechanisms, highlighting the need for awareness of these factors in early IA populations. Future longitudinal studies are essential to track the evolution of pain mechanisms over time and identify predictors of persistent pain, which may inform early interventions targeting central pain processes.

ACKNOWLEDGMENT

We thank Professor Thomas Graven Nielson for his invaluable advice and support in training and interpretation of cuff algometer data throughout the project.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Svensson B, Forslind K, Andersson M. Unacceptable pain in the BARFOT inception cohort of patients with rheumatoid arthritis: a long-term study. Scand J Rheumatol 2020;9(5):371–378.
2. Vergne-Salle P, Pouplin S, Trouvin AP, et al. The burden of pain in rheumatoid arthritis: impact of disease activity and psychological factors. Eur J Pain 2020;24(10):1979–1989.
3. Michaud K, Pope J, van de Laar M, et al. Systematic literature review of residual symptoms and an unmet need in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2021;73(11):1606–1616.
4. Conaghan PG, Alten R, Deodhar A, et al. Relationship of pain and fatigue with health-related quality of life and work in patients with psoriatic arthritis on TNFi: results of a multi-national real-world study. RMD Open 2020;6(2):6.
5. Chaabo K, Chan E, Garrood T, et al. Pain sensitisation and joint inflammation in patients with active rheumatoid arthritis. RMD Open 2024;10(1):e003784.
6. Rutter-Locher Z, Arumalla N, Norton S, et al. A systematic review and meta-analysis of questionnaires to screen for pain sensitisation and neuropathic-like pain in inflammatory arthritis. Semin Arthritis Rheum 2023;61:152207.

7. Walsh DA, McWilliams DF. Mechanisms, impact and management of pain in rheumatoid arthritis. Nat Rev Rheumatol 2014;10(10): 581–592.
8. Iyer P, Lee YC. Why it hurts: the mechanisms of pain in rheumatoid arthritis. Rheum Dis Clin North Am 2021;47(2):229–244.
9. Zhang A, Lee YC. Mechanisms for joint pain in rheumatoid arthritis (RA): from cytokines to central sensitization. Curr Osteoporos Rep 2018;16(5):603–610.
10. Lee YC, Nassikas NJ, Clauw DJ. The role of the central nervous system in the generation and maintenance of chronic pain in rheumatoid arthritis, osteoarthritis and fibromyalgia. Arthritis Res Ther 2011; 13(2):211.
11. Lee YC, Lu B, Edwards RR, et al. The role of sleep problems in central pain processing in rheumatoid arthritis. Arthritis Rheum 2013;65(1): 59–68.
12. Arant KR, Katz JN, Neogi T. Quantitative sensory testing: identifying pain characteristics in patients with osteoarthritis. Osteoarthritis Cartilage 2022;30(1):17–31.
13. Meeus M, Vervisch S, De Clerck LS, et al. Central sensitization in patients with rheumatoid arthritis: a systematic literature review. Semin Arthritis Rheum 2012;41(4):556–567.
14. Trouvin AP, Attal N, Perrot S. Assessing central sensitization with quantitative sensory testing in inflammatory rheumatic diseases: a systematic review. Joint Bone Spine 2022;89(5):105399.
15. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33(5):337–343.
16. Gossec L, Paternotte S, Aanerud GJ, et al. Finalisation and validation of the rheumatoid arthritis impact of disease score, a patient-derived composite measure of impact of rheumatoid arthritis: a EULAR initiative. Ann Rheum Dis 2011;70(6):935–942.
17. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258–266.
18. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–613.
19. Spitzer RL, Kroenke K, Williams JBW, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092–1097.
20. Wolfe F, Clauw DJ, Fitzcharles MA, et al. 2016 revisions to the 2010/2011 fibromyalgia diagnostic criteria. Semin Arthritis Rheum 2016;46(3):319–329.
21. Freynhagen R, Baron R, Gockel U, et al. painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. Curr Med Res Opin 2006;22(10):1911–1920.
22. Ahmed S, Magan T, Vargas M, et al. Patient reported pain by the painDETECT questionnaire reveals multimodal elements to pain perception in rheumatoid arthritis [abstract]. Arthritis Rheumatol 2014;66(suppl 10):S1297.
23. Rifbjerg-Madsen S, Christensen AW, Boesen M, et al. The course of pain hypersensitivity according to painDETECT in patients with rheumatoid arthritis initiating treatment: results from the prospective FRAME-cohort study. Arthritis Res Ther 2018;20(1):105.
24. Rifbjerg-Madsen S, Christensen AW, Christensen R, et al. Pain and pain mechanisms in patients with inflammatory arthritis: a Danish nationwide cross-sectional DANBIO registry survey. PLoS One 2017;12(7):e0180014.
25. D'Agostino MA, Terslev L, Aegerter P, et al. Scoring ultrasound synovitis in rheumatoid arthritis: a EULAR-OMERACT ultrasound taskforce-part 1: definition and development of a standardised, consensus-based scoring system. RMD Open 2017;3(1):e000428.
26. Rolke R, Magerl W, Campbell KA, et al. Quantitative sensory testing: a comprehensive protocol for clinical trials. Eur J Pain 2006;10(1): 77–88.
27. Magerl W, Krumova EK, Baron R, et al. Reference data for quantitative sensory testing (QST): refined stratification for age and a novel method for statistical comparison of group data. Pain 2010;151(3): 598–605.
28. Graven-Nielsen T, Arendt-Nielsen L. Assessment of mechanisms in localized and widespread musculoskeletal pain. Nat Rev Rheumatol 2010;6(10):599–606.
29. Vaegter HB, Graven-Nielsen T. Pain modulatory phenotypes differentiate subgroups with different clinical and experimental pain sensitivity. Pain 2016;157(7):1480–1488.
30. Graven-Nielsen T, Izumi M, Petersen KK, et al. User-independent assessment of conditioning pain modulation by cuff pressure algometry. Eur J Pain 2017;21(3):552–561.
31. Vollert J, Maier C, Attal N, et al. Stratifying patients with peripheral neuropathic pain based on sensory profiles: algorithm and sample size recommendations. Pain 2017;158(8):1446–1455.
32. Rolke R, Baron R, Maier C, et al. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. Pain 2006;123(3):231–243.
33. Norton S, Ellis B, Santana Suárez B, et al. Validation of the Musculoskeletal Health Questionnaire in inflammatory arthritis: a psychometric evaluation. Rheumatology (Oxford) 2019;58(1):45–51.
34. Pollard LC, Ibrahim F, Choy EH, et al. Pain thresholds in rheumatoid arthritis: the effect of tender point counts and disease duration. J Rheumatol 2012;39(1):28–31.
35. Tabachnick BG, Fidell, LS. Using Multivariate Statistics. 4th ed. Allyn & Bacon; 2001.
36. Altawil R, Saevarsdottir S, Wedrén S, et al. Remaining pain in early rheumatoid arthritis patients treated with methotrexate. Arthritis Care Res (Hoboken) 2016;68(8):1061–1068.
37. Fragoulis GE, Cavanagh J, Tindell A, et al. Depression and anxiety in an early rheumatoid arthritis inception cohort. Associations with demographic, socioeconomic and disease features. RMD Open 2020;6(3):e001376.
38. Heidari F, Afshari M, Moosazadeh M. Prevalence of fibromyalgia in general population and patients, a systematic review and meta-analysis. Rheumatol Int 2017;37(9):1527–1539.
39. Lee YC, Lu B, Boire G, et al. Incidence and predictors of secondary fibromyalgia in an early arthritis cohort. Ann Rheum Dis 2013;72(6): 949–954.
40. Yates M, Ledingham JM, Hatcher PA, et al. Disease activity and its predictors in early inflammatory arthritis: findings from a national cohort. Rheumatology (Oxford) 2021;60(10):4811–4820.
41. Zhao L, Galloway J, Ledingham J, et al. Psychological distress over 12 months post-diagnosis in an early inflammatory arthritis cohort. Rheumatology (Oxford) 2025;64(5):2469–2478.
42. Lee YC, Bingham CO III, Edwards RR, et al. Association between pain sensitization and disease activity in patients with rheumatoid arthritis: a cross-sectional study. Arthritis Care Res (Hoboken) 2018;70(2): 197–204.
43. Rifbjerg-Madsen S, Wæhrens EE, Danneskiold-Samsøe B, et al. Psychometric properties of the painDETECT questionnaire in rheumatoid arthritis, psoriatic arthritis and spondyloarthritis: rasch analysis and test-retest reliability. Health Qual Life Outcomes 2017;15(1):110.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 1, January 2026, pp 164–174

LETTER

DOI 10.1002/acr.25559

Key considerations for acetaminophen safety in older adults with osteoarthritis: comment on the article by Kaur et al

To the Editor:

We read with great interest the article related to acetaminophen and its potential side effects in older adults.[1] However, there are some points to take into consideration.

First, there is a higher prevalence of comorbid conditions in individuals with osteoarthritis (OA). A meta-analysis by Swain et al[2] reported that 67% (95% confidence interval [CI] 57–74) of individuals with OA had at least one comorbidity, compared to 56% (95% CI 44–68) of those without OA. Among the most common chronic conditions in patients with OA, hypertension (HTN) was present in 50%, cardiovascular disease (CVD) was present in 35%, peptic ulcer disease was present in 16%, and diabetes was present in 14%. Moreover, Hall et al[3] identified that individuals with OA are twice as likely to develop CVD. Therefore, despite excluding patients from the study with diagnosed conditions, subclinical cases may persist and bias the study's conclusions.[1]

Secondly, although acetaminophen is not contraindicated in patients with OA with HTN, there are conflicting data on its effect on CVD. Fulton et al[4] conducted a study including 10,878 hypertensive individuals exposed to acetaminophen during a 10-year period. There was no relationship between risk of myocardial infarction (hazard ratio [HR] 0.98; 95% CI 0.76–1.27) or any cardiovascular event (HR 1.17; 95% CI 0.99–1.37) and acetaminophen exposure. However, Roberts et al[5] revealed an increased risk ratio for all cardiovascular adverse events, ranging from 1.19 (95% CI 0.81–1.75) to 1.68 (95% CI 1.10–2.57). They additionally identified an increasing odds ratio for a ≥30% decline in the estimated glomerular filtration rate, ranging from 1.40 (95% CI 0.79–2.48) to 2.19 (95% CI 1.40–3.43). In summary, despite potential bias, the dose–response trend suggests significant paracetamol toxicity, especially at high doses. This risk is greater in older adults because of age-related metabolic changes, warranting a lower maximum daily dose of 2 to 3 g.

Lastly, both acetaminophen and nonsteroidal anti-inflammatory drugs (NSAIDs) are widely accessible as over-the-counter pain medications. This easy access may lead to exposure misclassification by not taking into consideration self-medication. In fact, Cavagna et al found that among 506 patients with OA observed for nine months, 240 (47%) used additional NSAIDs, either prescribed or self-administered. About 20% cited incomplete pain control.[6] Therefore, future studies should incorporate pharmacy data or self-reported medication use to improve accuracy.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25559.

Fernando Gallardo-Landauro, MBBS
fernando.gallardo@upch.pe
Germán Delgado de la Flor-Aquino, MBBS
Luis A. Fernández-Cornejo, MBBS
Universidad Peruana Cayetano Heredia Facultad de Medicina Alberto Hurtado
Lima, Peru

1. Kaur J, Nakafero G, Abhishek A, et al. Incidence of side effects associated with acetaminophen in people aged 65 years or more: a prospective cohort study using data from the Clinical Practice Research Datalink. Arthritis Care Res (Hoboken) 2025;77(5):666–675.
2. Swain S, Sarmanova A, Coupland C, et al. Comorbidities in osteoarthritis: a systematic review and meta-analysis of observational studies. Arthritis Care Res (Hoboken) 2020;72(7):991–1000.
3. Hall AJ, Stubbs B, Mamas MA, et al. Association between osteoarthritis and cardiovascular disease: systematic review and meta-analysis. Eur J Prev Cardiol 2016;23(9):938–946.
4. Fulton RL, Walters MR, Morton R, et al. Acetaminophen use and risk of myocardial infarction and stroke in a hypertensive cohort. Hypertension 2015;65(5):1008–1014.
5. Roberts E, Delgado Nunes V, Buckner S, et al. Paracetamol: not as safe as we thought? A systematic literature review of observational studies. Ann Rheum Dis 2016;75(3):552–559.
6. Cavagna L, Caporali R, Trifiro G, et al. Overuse of prescription and OTC non-steroidal anti-inflammatory drugs in patients with rheumatoid arthritis and osteoarthritis. Int J Immunopathol Pharmacol 2013;26(1):279–281.

DOI 10.1002/acr.25560

Reply

To the Editor:

We thank Gallardo-Landauro et al for their interest in this study. We conducted a retrospective cohort study and reported that acetaminophen prescription was associated with a significantly higher risk of gastrointestinal, cardiovascular, and renal adverse effects in adults aged ≥65 years. We acknowledge the concerns raised by Gallardo-Landauro et al regarding the potential bias introduced by "subclinical cases." To paraphrase Gallardo-Landauro et al, we presume by this that they mean people with an outcome that is at a mild and/or early phase and

without a diagnosis recorded. To minimize any issues from reverse causality, we used landmark analysis, inverse probability treatment weighting, and propensity score matching to balance the acetaminophen-exposed and unexposed groups, incorporating multiple covariates.[1] However, as mentioned in the limitations, residual confounding remains an issue.

We acknowledge that some studies find no strong link between acetaminophen and cardiovascular risks, whereas others[2–4] suggest potential harm at high doses, as already mentioned in the original article.[1] We had already covered in the study that over-the-counter (OTC) medications play a crucial role in self-care, but their use is often not recorded in electronic health records, posing a challenge in studies like this. To minimize this limitation, our study[1] focused on individuals aged ≥65 years who were less likely to purchase acetaminophen independently because of free prescriptions in this age group. However, OTC use may still have affected both exposure and nonexposure groups, likely leading to an underestimation rather than an overestimation of the hazard ratio.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25560.

Jaspreet Kaur, BAMS, MPH, PG Diploma, PhD
jaspreet.kaur1@nottingham.ac.uk
Abhishek Abhishek, MBBS, MD (Medicine), FRCP (UK), PhD
University of Nottingham
Nottingham, United Kingdom
Georgina Nakafero, BScN, MPH, PhD
University of Nottingham and National Institute for Health and Care Research Biological
Research Centre
Nottingham, United Kingdom
Christian Mallen, BMedSci, BMBS, MMedSci, MPhil, PhD, FRCGP, FFCI, FFPH
Keele University
Keele, United Kingdom
Michael Doherty, MA, MD, FRCP
Weiya Zhang, BMed, MMed, PhD
University of Nottingham and Pain Centre Versus Arthritis
Nottingham, United Kingdom

1. Kaur J, Nakafero G, Abhishek A, et al. Incidence of side effects associated with acetaminophen in people aged 65 years or more: a prospective cohort study using data from the Clinical Practice Research Datalink. Arthritis Care Res (Hoboken) 2025;77(5):666–675.
2. Spence JD, Grosser T, FitzGerald GA. Acetaminophen, nonsteroidal anti-inflammatory drugs, and hypertension. Hypertension 2022;79(9):1922–1926.
3. Chan AT, Manson JE, Albert CM, et al. Nonsteroidal antiinflammatory drugs, acetaminophen, and the risk of cardiovascular events. Circulation 2006;113(12):1578–1587.
4. Roberts E, Delgado Nunes V, Buckner S, et al. Paracetamol: not as safe as we thought? A systematic literature review of observational studies. Ann Rheum Dis 2016;75(3):552–559.

DOI 10.1002/acr.25586

Concerns regarding pentraxin-3 specificity and receiver operating characteristic curve interpretation in the study on endothelial biomarkers of systemic sclerosis–associated pulmonary hypertension: comment on the article by Lammi et al

To the Editor:

We read with great interest the recent article by Lammi et al on endothelial biomarkers in systemic sclerosis–associated pulmonary hypertension (SSc-PH).[1] The effort to refine noninvasive risk stratification in SSc-PH is commendable, especially the focus on pentraxin-3 (PTX-3) as a potential discriminative biomarker. However, we wish to raise several concerns regarding the interpretation of their findings and the article overall.

Firstly, although the title suggests a biomarker discovery study, it does not clarify the actual study design, intent, or clinical setting. Additionally, the stated objective in the abstract and introduction is loosely defined, and the role of individual biomarkers is not clearly delineated. Neither asymmetric dimethylarginine nor soluble endoglin showed meaningful discriminatory value across groups, and there was no correlation among the three endothelial biomarkers. This weakens the rationale for presenting them as a unified "endothelial panel" and raises questions about the primary focus of the analysis. Also, the cohort of interest was selected based on patients who had blood samples collected (the group more likely to have PH), and right-sided heart catheterization was not performed in all high-risk patients. Although the authors acknowledge a few of these limitations, the results may not apply to broader community-based or early SSc populations.

Secondly, the internal biologic consistency of PTX-3 behavior across the study groups raises concern. The study reports higher PTX-3 values in healthy controls (median ∼77 pg/mL), patients at a high risk of SSc (median ∼68 pg/mL), and patients with SSc-PH (median ∼83.5 pg/mL) yet strikingly low levels in the low-risk SSc group (median 27 pg/mL). This is counterintuitive because PTX-3 is an acute-phase reactant, and levels are generally elevated in inflammatory and vascular disease states. It is difficult to explain why patients with SSc (a known vasculopathic condition), even at low risk for PH, would have lower PTX-3 levels than healthy individuals, unless influenced by sampling or technical artifacts. Although the authors speculate that lower PTX-3 levels may reflect improved endothelial regenerative capacity, this remains speculative and lacks supportive functional evidence. Prior studies have consistently shown elevated PTX-3 levels in SSc-PH and other vascular disease states.[2] Although the authors report a statistically significant difference in PTX-3 levels between low- and high-risk groups, the interquartile ranges (IQRs) show near overlap, casting doubt on the clinical relevance of this difference. This raises doubts about the real-world utility of PTX-3 as

a standalone discriminator, considering the small sample size in the low-risk group.

Also, the authors propose that a PTX-3 cutoff of 51.5 pg/mL yields "100% specificity" for identifying patients with SSc at low risk for PH. However, their own data report an IQR of 48.5 to 104.5 pg/mL for the high-risk group, indicating that the proposed cutoff lies within the IQR of high-risk individuals. This directly challenges the assertion of perfect specificity and raises concerns about misclassification risk. The authors do not report whether this threshold was derived using an objective method such as the Youden index, nor do they provide a table of sensitivity and specificity values for alternate cutoffs. Moreover, this conclusion is based on a small sample size (n = 10) in the low-risk group, which renders the receiver operating characteristic curve–derived specificity statistically unstable and potentially nongeneralizable.[3] So even if none of them had a PTX-3 value above 51.5 pg/mL and this produces an apparent 100% specificity (10 of 10), such a value derived from such a small subgroup is statistically fragile and prone to overestimation.

Lastly, the lack of a validation cohort further limits the robustness of the conclusions, particularly in light of these inconsistencies. Although the authors mention the small sample size and lack of a validation cohort in their discussion on limitations, they do also mention that they were able to stratify the patients based on their risk profile for PH as their merit. We agree that PTX-3 warrants further exploration as a biomarker in SSc-PH but urge caution in interpreting these findings. Larger prospectively validated cohorts are essential to clarify the diagnostic and prognostic utility of PTX-3 in this.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25586.

Ashwant Kumar, MBBS, MD
Sabarinath Mahadevan, MBBS, MD, DM, MRCP, SCE
Rheumatology
msabari1987@gmail.com
Institute of Rheumatology, Madras Medical College
Chennai, India

1. Lammi MR, Kolstad KD, Saketkoo LA, et al. Endothelial biomarkers of systemic sclerosis–associated pulmonary hypertension. Arthritis Care Res (Hoboken) 2025;77(2):266–272.
2. Shirai Y, Okazaki Y, Inoue Y, et al. Elevated levels of pentraxin 3 in systemic sclerosis: associations with vascular manifestations and defective vasculogenesis. Arthritis Rheumatol 2015;67(2):498–507.
3. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med 2013; 4(2):627–635.

DOI 10.1002/acr.25589

Reply

To the Editor:

We thank Drs Kumar and Mahadevan for their interest in our publication "Endothelial biomarkers of systemic sclerosis–associated pulmonary hypertension."[1] They raise some important questions about our study methodology and interpretation, and we appreciate the ability to address them in turn. Drs Kumar and Mahadevan are correct that right-sided heart catheterization was not performed in all patients with systemic sclerosis (SSc) with high risk for pulmonary hypertension (PH) features, which may have led to a misclassification of study patients. Unfortunately, this was a consequence of the Pulmonary Hypertension Assessment and Recognitions of Outcomes in Scleroderma (PHAROS) registry protocol, but we do not believe that this is likely to have significantly affected the biomarker results because there were no suggestions of difference in pentraxin-3 (PTX-3), soluble endoglin, or asymmetric dimethylarginine levels between patients with SSc-PH and the patients with SSc with high risk for PH. There were notable differences between patients in PHAROS who had blood samples collected and the patients without available blood samples, which we clearly outline in Supplemental Table 1. In reviewing our data for accuracy, we discovered an error in the data analysis of PH status stratified by availability of blood samples. There was no significant difference in PH status based on whether patients did or did not have blood samples available for analysis (P = 0.08; blood sample available: high risk for PH n = 44 [37%], PH n = 74 [63%]; blood sample not available: high risk for PH n = 203 [47%], PH n = 229 [53%]). This has been updated in the Supplementary Materials available online.

Drs Kumar and Mahadevan describe our finding that PTX-3 levels were lowest in the low risk for PH group, even lower than those in healthy controls, as being counterintuitive. We fully agree with this assessment and stated that our hypothesis was that the endothelial biomarkers (eg, PTX-3) would be higher in patients with high risk for PH compared to patients with low risk for PH. However, the statistical testing suggests that this difference is unlikely to be due to chance alone. Regarding the method for selecting the reported PTX-3 cutoff of 51.5 pg/mL, this is the cutoff that yielded the highest Youden index of 0.68, which describes the value of the biomarker with the highest combination of sensitivity and specificity. We refer the interested reader to Supplemental Table 2 of our original article for alternative PTX-3 cutoffs with their associated sensitivities and specificities. Drs Kumar and Mahadevan describe our posited explanation for the lower values of PTX-3 seen in patients with low risk for SSc as speculative, which is the same phrase that we used in our Discussion to describe this potential mechanism. We also discussed the fact that our findings differ from prior reports of elevated PTX-3 levels in patients with SSc with PH,[2] but without careful characterization of the risk profile contained within the "non-PH" group, findings such as this may have been missed. We would also like to highlight that PTX-3 levels were higher in patients with SSc with precapillary PH compared to those with postcapillary PH, further suggesting

that differences may exist within the spectrum of SSc and its associated pulmonary vascular disease. Lastly, we agree that our findings are limited due to the lack of a validation cohort, which is why we called for replication within a larger SSc cohort in our conclusions. We hope that our study will help stimulate further research into PTX-3 as a possible biomarker of SSc phenotypes.

Matthew Lammi, MD, MSCR
mlammi@lsuhsc.edu
The Johns Hopkins University School of Medicine
Baltimore, MD
Kathleen D. Kolstad, MD, PhD
University of California, Los Angeles
Lesley Ann Saketkoo, MD, MPH
Tulane University School of Medicine
New Orleans, LA
Avani Khatri, PhD
Paul J. Utz, PhD
Lorinda Chung, MD, MS
Stanford University School of Medicine
Stanford, CA
Virginia Steen, MD
Georgetown University Medical Center
Washington, DC

1. Lammi MR, Kolstad KD, Saketkoo LA, et al. Endothelial biomarkers of systemic sclerosis–associated pulmonary hypertension. Arthritis Care Res (Hoboken) 2025;77(2):266–272.
2. Shirai Y, Okazaki Y, Inoue Y, et al. Elevated levels of pentraxin 3 in systemic sclerosis: associations with vascular manifestations and defective vasculogenesis. Arthritis Rheumatol 2015;67(2):498–507.

DOI 10.1002/acr.25587

Risk of hepatotoxicity in use of febuxostat or benzbromarone: comment on the article by Sun et al

To the Editor:

Sun et al[1] recently reported a significantly increased risk of hepatotoxicity in febuxostat users compared with benzbromarone users among patients with gout (39.6 vs 16.8 events per 1,000 person-years; adjusted hazard ratio 2.75; 95% confidence interval [CI] 1.28–5.91).[1] Their findings are notable because they use Common Terminology Criteria for Adverse Events (CTCAE) v5.0 to define hepatotoxicity, highlighting not just biochemical elevations but clinically relevant thresholds. In contrast, the meta-analysis by Wu et al,[2] which included studies from Chinese and Japanese populations, found no significant difference in hepatic or renal safety between febuxostat and benzbromarone but relied primarily on mean alanine aminotransferase (ALT) and aspartate aminotransferase (AST) values without applying CTCAE-based grading. Sun et al used a more detailed and clinically relevant grading system to assess hepatotoxicity severity. In contrast, Wu et al focused on average enzyme levels to compare liver safety between febuxostat and benzbromarone. Using mean values can mask the presence of severe cases of hepatotoxicity because it averages out the enzyme levels, potentially diluting the impact of high values seen in some patients. That is, high enzyme levels in some patients might be less noticeable because they are combined with lower levels from other patients, resulting in an average that does not fully reflect the severity of the high values. Therefore, although the study by Wu et al provides a general overview of liver enzyme levels, it does not capture the clinical significance or severity of liver damage as effectively as the CTCAE grading system used by Sun et al.

These divergent conclusions reflect differences in outcome definitions, study designs, and, possibly, underlying comorbidity patterns. For example, the study by Sun et al may have included a higher prevalence of patients with metabolic syndrome and chronic viral hepatitis,[1] both of which increase susceptibility to liver injury. Additionally, febuxostat was associated with more ALT and AST elevations.[1] Wu et al demonstrated that benzbromarone led to a modest but statistically significant reduction in serum uric acid levels compared to febuxostat, with a weighted mean difference of -0.60 mg/dL (95% CI -1.09 to -0.12).[2] However, the clinical relevance of this small average difference may be limited. This highlights that statistical significance does not always equate to clinical impact, especially when the difference (-0.60 mg/dL) may not be substantial enough to justify choosing one drug over the other based solely on efficacy.

Although concerns over hepatotoxicity led to benzbromarone's withdrawal from many markets in the early 2000s—after case reports such as that by van der Klauw et al[3]—later reviews questioned whether this decision was justified.[4] In particular, benzbromarone remains in use not only in Taiwan and Japan but also in countries across Latin America, the Middle East, and parts of Europe and Africa.[5] Benzbromarone's continued clinical use—particularly in patients with urate underexcretion—suggests a favorable benefit–risk profile when appropriately selected and monitored. Furthermore, recent observational data have associated benzbromarone use with a lower risk of developing chronic kidney disease compared to allopurinol among adults with asymptomatic hyperuricemia, reinforcing its nephroprotective potential in real-world settings.[6]

Based on current data, benzbromarone may be considered superior in urate-lowering efficacy and remains a valuable therapeutic option for patients in whom effective serum uric acid reduction is the primary goal. Febuxostat remains important where

access to benzbromarone is limited or contraindicated. A re-evaluation of the safety profiles of both febuxostat and benzbromarone in future research may support broader access and more individualized treatment strategies.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25587.

Shih-Wei Lai, MD, PhD
School of Medicine, China Medical University and Department of Family Medicine, China Medical University Hospital
Taichung, Taiwan
Kuan-Fu Liao, MD, PhD
kuanfuliaog@gmail.com
Division of Hepatogastroenterology, Department of Internal Medicine
Taichung Tzu Chi Hospital
Taichung, Taiwan

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score–matched cohort study. Arthritis Care Res (Hoboken) 2025;77(9):1149–1156.
2. Wu F, Chen L, Du Y. Comparison of the efficacy and safety of benzbromarone and febuxostat in gout and hyperuricemia: a systematic review and meta-analysis. Clin Rheumatol 2024;43(5):1745–1754.
3. van der Klauw MM, Houtman PM, Stricker BH, et al. Hepatic injury caused by benzbromarone. J Hepatol 1994;20(3):376–379.
4. Lee MH, Graham GG, Williams KM, et al. A benefit-risk assessment of benzbromarone in the treatment of gout. Was its withdrawal from the market in the best interest of patients? Drug Saf 2008;31(8):643–665.
5. Drugs.com. Benzbromarone (International). Accessed April 1, 2025. https://www.drugs.com/international/benzbromarone.html
6. Lai SW, Liao KF, Kuo YH, et al. Comparison of benzbromarone and allopurinol on the risk of chronic kidney disease in people with asymptomatic hyperuricemia. Eur J Intern Med 2023;113:91–97.

DOI 10.1002/acr.25592

Reply

To the Editor:

We thank Drs Lai and Liao for their thoughts on the safety-risk profile of benzbromarone and febuxostat. As demonstrated in our study,[1] febuxostat was more likely to cause mild-to-moderate perturbation of liver function in patients with gout than benzbromarone, which was consistent with prior findings.[2] Notably, severe hepatotoxicity remains rare with either benzbromarone or febuxostat in clinical practice. Additional results on alanine aminotransferase (ALT) or aspartate aminotransferase (AST) fluctuation trends by a generalized additive mixed model further indicated that no statistically significant differences between those started on febuxostat and those started on benzbromarone (Figure 1). Although some randomized controlled trials routinely document ALT and AST during treatment, Common Terminology Criteria for Adverse Events grading for adverse events or more severe episodes was usually omitted. Therefore, as Drs Lai and Liao pointed out, adopting clinically relevant thresholds for outcome assessment may prove crucial.

Liver function abnormalities occur in 20% to 40% of patients with gout, which was similar in our cohort (~31%). Pre-existing liver disease requires one to be more cautious when choosing between urate-lowering drugs. Our analysis identified baseline elevated ALT and AST levels as independent risk factors for drug-induced hepatotoxicity in both those started on febuxostat and those started on benzbromarone. This aligns with some guidelines[3] recommending monitoring liver function during urate-lowering therapy (ULT).

Furthermore, according to Lai et al, the incidence rate of chronic kidney disease (CKD) was lower in the benzbromarone group than in the allopurinol group in persons with asymptomatic hyperuricemia.[4] Potential renal advantages of benzbromarone could be attributable to its uricosuric mechanism preserving renal tubular function. Recently, a post hoc analysis of the CARES trial demonstrated that ULT with either allopurinol or febuxostat delays CKD progression.[5] A well-designed study comparing the renal outcomes of benzbromarone versus febuxostat in people with gout is warranted.

Benzbromarone marketing remains markedly restricted geographically because of uncommon but potentially lethal, mitochondrial damage-mediated hepatotoxicity.[6] Hepatic safety, particularly in those with pre-existing liver disease, is an important consideration in prescribing either of these medications. Certain identifiable susceptibility factors for hepatotoxicity can help in designing benzbromarone analogues with potential safety advantages. Significantly, the current drug development pipeline of effective, uricosuric benzbromarone analogues with highly potent urate-lowering activity and drug design features to eliminate risk of severe hepatotoxicity should be beneficial in the long run.

Supported by the Shandong Provincial Natural Science Foundation (ZR2023MH213).

Lingling Cui, PhD
Wenyan Sun, PhD
Changgui Li, MD
changguili@vip.163.com
Shandong Provincial Clinical Research Center for Immune Diseases and Gout, Shandong Provincial Clinical Medical Center for Gout, the Affiliated Hospital of Qingdao University
Qingdao, China

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score–matched cohort study. Arthritis Care Res (Hoboken) 2025;77(9):1149–1156.
2. Bohm M, Vuppalanchi R, Chalasani N; Drug-Induced Liver Injury Network (DILIN). Febuxostat-induced acute liver injury. Hepatology 2016;63(3):1047–1049.

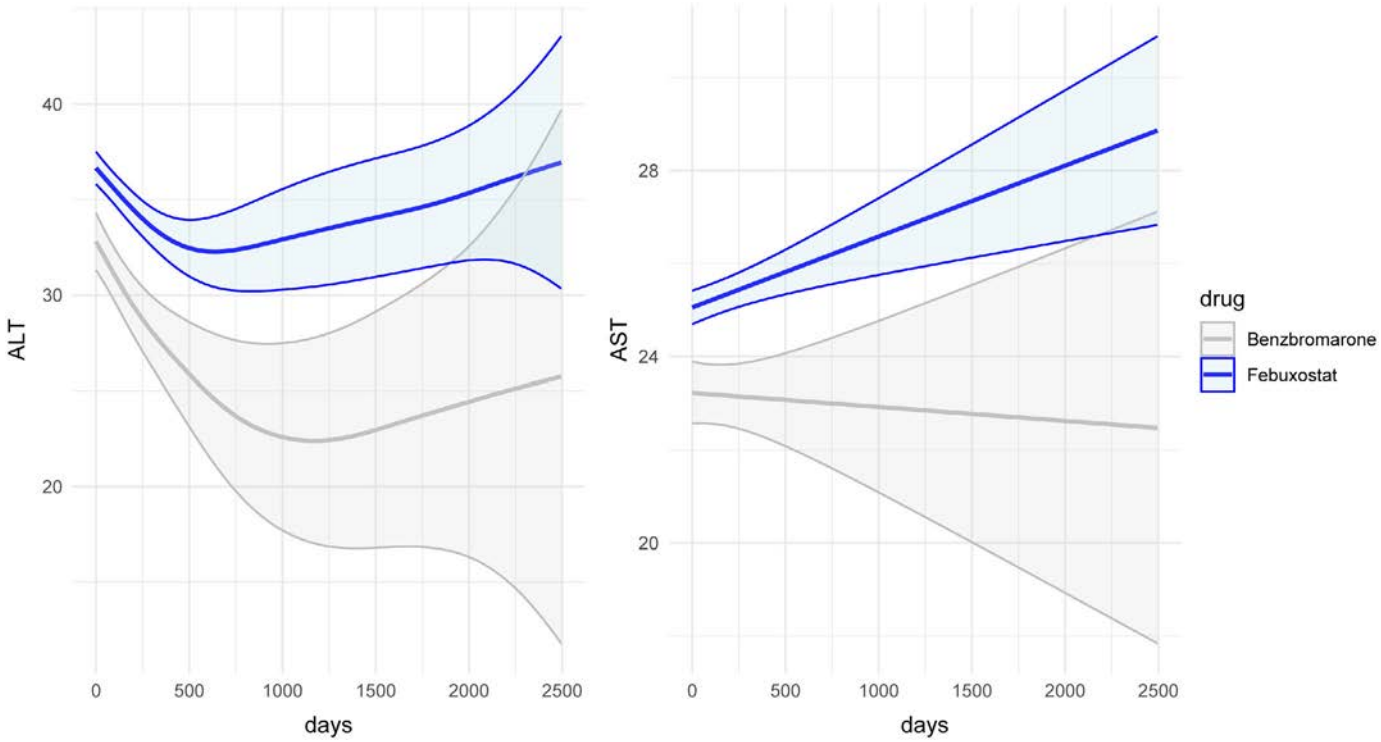


Figure 1. Changes in AST and ALT levels over time in the febuxostat group and benzbromarone group. ALT, alanine aminotransferase; AST, aspartate aminotransferase.

3. Yu KH, Chen DY, Chen JH, et al. Management of gout and hyperuricemia: Multidisciplinary consensus in Taiwan. Int J Rheum Dis 2018;21(4):772–787.
4. Lai SW, Liao KF, Kuo YH, et al. Comparison of benzbromarone and allopurinol on the risk of chronic kidney disease in people with asymptomatic hyperuricemia. Eur J Intern Med 2023;113:91–97.
5. Ghang B, Park J, Lee JS, et al. Post-hoc analysis of the CARES trial suggests delayed progression of chronic kidney disease in patients with gout during urate-lowering therapy. Kidney Int 2025;107(3):521–529.
6. Terkeltaub R. Emerging urate-lowering drugs and pharmacologic treatment strategies for gout: a narrative review. Drugs 2023;83(16):1501–1521.

DOI 10.1002/acr.25591

Febuxostat use is associated with a significantly greater risk of mild to moderate perturbation of liver function compared to benzbromarone in patients with gout: comment on the article by Sun et al

To the Editor:

We read with interest the recent publication by Sun and colleagues on the risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone.[1] The authors concluded that febuxostat use is associated with a significantly greater risk of mild to moderate perturbation of liver function compared to benzbromarone in patients with gout.[1] We support and appreciate the authors' work and agree with their conclusions but have some concerns about some of the details in the article.

First, there is insufficient detail regarding the liver function status at baseline. The study only recorded alanine transaminase and aspartate transferase levels at baseline but did not conduct a comprehensive assessment of liver function, such as liver ultrasound or elastography.[1] These tests could provide more detailed information on liver structure and function, helping to more accurately assess the baseline liver function status of the patients. Furthermore, the study did not provide a detailed classification of patients with chronic liver disease, such as distinguishing between fatty liver, hepatitis, cirrhosis, and other types. The liver toxicity risks in these patients with different types of liver disease may vary, and the lack of detailed classification may affect the accurate assessment of liver toxicity risk.

Second, lifestyle factors (such as alcohol consumption, smoking, and exercise habits) could significantly impact liver function. For example, alcohol consumption is a known hepatotoxic factor and may interact with the hepatotoxicity of drugs.[2,3] However, the study did not record the frequency, amount, or type of alcohol consumed by the patients in detail, making it impossible to assess the impact of these factors on liver toxicity.[1] Dietary habits (such as high-fat or high-purine diets) could also influence liver function and drug metabolism. For example, a high-fat diet may lead to fatty liver, increasing the risk of drug-induced liver injury. The study did not provide a detailed assessment of the patient's dietary habits, and thus the potential influence of these factors on the study results cannot be ruled out.

Third, the inconsistency in follow-up time may be one of the overlooked limitations of this study. The average follow-up time for the febuxostat group (1.34 years) was longer than that for the benzbromarone group (1.09 years), and this discrepancy in follow-up time could influence the comparison of results.[1] For example, a longer follow-up time may result in more liver toxicity events being observed, potentially overestimating the liver toxicity risk of febuxostat.

Fourth, although this study used propensity score matching and inverse probability of treatment weighting to control for potential confounding factors, these methods may not completely eliminate the effects of all confounders.[1] For instance, the study did not consider other medications that patients may have been taking concurrently (such as acetaminophen, antibiotics, etc) that could also impact liver function.

Finally, the metabolism and clearance processes of febuxostat and benzbromarone may vary because of individual differences.[1] For example, some patients may experience differences in the metabolic rate and clearance efficiency of the drugs because of genetic factors or variations in liver and kidney function, which could affect liver toxicity risk. However, the study did not conduct dynamic monitoring of drug metabolism and clearance, so it was unable to assess the impact of these factors on liver toxicity.

In conclusion, before these issues are clarified, this study's findings should be interpreted cautiously.

We would like to thank the members and staff of the Department of Rheumatology and Immunology of The Second Affiliated Hospital of Soochow University who contributed to this manuscript.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25591.

Gang Wang, MD
Zhichun Liu, MD
zcliurheu@suda.edu.cn
The Second Affiliated Hospital of Soochow University
Suzhou, China

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score-matched cohort study. Arthritis Care Res (Hoboken) 2025;77(9):1149–1156.
2. Rusyn I, Bataller R. Alcohol and toxicity. J Hepatol 2013;59(2):387–388.
3. Sheron N. Alcohol and liver disease in Europe--simple measures have the potential to prevent tens of thousands of premature deaths. J Hepatol 2016;64(4):957–967.

DOI 10.1002/acr.25590

Reply

To the Editor:

Thanks to Drs Wang and Liu for raising the critical points regarding hepatic effects of benzbromarone and febuxostat in our study. We acknowledge that although propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) were used to control for confounders, certain limitations remain. First, comprehensive baseline liver function evaluations, such as liver ultrasound or elastography, will provide more detailed structural and functional information, particularly for detecting steatosis or fibrosis. In our study, patients with liver disease were defined as those with hepatosteatosis or abnormal liver function test results (aspartate transferase [AST] or alanine transaminase [ALT] level more than the upper limit of normal line) at baseline.[1] This approach is consistent with similar observational studies prioritizing biochemical markers (ALT and AST) as primary end points. However, it should be noted that no patients with cirrhosis or fibrosis at baseline were included in our cohort at baseline. Emerging evidence suggests differential drug sensitivity across liver disease subtypes,[2] a risk factor we plan to prioritize in subsequent research designs.

Regarding alcohol consumption as a documented hepatotoxic agent that may synergize with drug metabolism pathways,[3] we recognize that incomplete data on alcohol consumption patterns and smoking habits (including frequency, quantity, and type) limit our ability to fully interpret hepatotoxicity risks. Our subgroup analysis did not reveal significant interaction effects between drinkers and nondrinkers or between smokers and nonsmokers.[1] Similarly, dietary factors such as high-fat diets promoting hepatic steatosis were not accounted for in our analysis. Although these limitations are acknowledged in our study, future pharmacovigilance research should aim to quantify the effect of lifestyle exposures.

As to the discrepancy in follow-up times (1.34 vs 1.09 years) that may introduce bias in outcome comparisons, in the sensitivity analysis, follow-up time was limited to 180 or 365 days in the matched cohort, and febuxostat use remained associated with more hepatotoxicity episodes compared to benzbromarone. Although PSM and IPTW effectively balance observed covariates between treatment groups, they cannot account for unmeasured confounders, such as concurrent medications, drug metabolism, and clearance. We agree that future well-designed studies are required to compare the hepatic effects of benzbromarone and febuxostat in patients with gout.

Supported by the Shandong Provincial Natural Science Foundation (grant ZR2023QH147).

Wenyan Sun, PhD
Lingling Cui, PhD
Changgui Li, MD
changguili@vip.163.com
Shandong Provincial Clinical Research Center for Immune Diseases and Gout, Shandong Provincial Clinical Medical Center for Gout, the Affiliated Hospital of Qingdao University
Qingdao, China

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score-matched cohort study. Arthritis Care Res (Hoboken) 2025;77(9):1149–1156.

2. Tang LWT, Varma MVS. Hepatic impairment and the differential effects on drug clearance mechanisms: analysis of pharmacokinetic changes in disease state. Clin Pharmacol Ther Published online March 6, 2025. https://doi.org/10.1002/cpt.3643
3. Humphreys JH, Warner A, Costello R, et al. Quantifying the hepatotoxic risk of alcohol consumption in patients with rheumatoid arthritis taking methotrexate. Ann Rheum Dis 2017;76(9):1509–1514.

DOI 10.1002/acr.25606

Validation of lung ultrasound interpretation criteria for ILD in systemic sclerosis and inflammatory myopathy: comment on the article by Fairchild et al

To the Editor:

We have read with great interest the article "Validation of Lung Ultrasound Interpretation Criteria for Interstitial Lung Disease in Systemic Sclerosis and Inflammatory Myopathy."[1] Interstitial lung disease (ILD) represents a significant and potentially progressive complication in both systemic sclerosis (SSc) and inflammatory myopathies, and early diagnosis is key to improving patient outcomes.[2] Lung ultrasound (LUS) is a relatively harmless, more accessible alternative to high-resolution computed tomography (HRCT), allowing for repeated assessments during clinical visits. This facilitates real-time monitoring, potentially improving early detection and evaluation of treatment response.

This study represents an important step toward that goal, and we commend the authors for their valuable contribution. However, we would like to comment on several of their findings.

The authors excluded B-lines as a criterion for ILD, citing their variable presence and the potential for marked reduction due to image-processing algorithms designed to eliminate artifacts. We believe this approach may have limitations in terms of diagnosis, prognosis, and treatment planning.

Regarding diagnosis, previous studies have shown that the total number of B-lines can significantly distinguish patients with ILD from those without in the context of SSc.[3] Furthermore, posterior and posterolateral B-lines were able to differentiate limited ILD from no ILD, which aligns with the natural history of ILD in SSc—typically beginning in the posterobasal regions and progressing toward anterior and proximal segments. This suggests that ultrasound may detect early, subclinical interstitial involvement.[3] In addition, some studies have indicated that pleural irregularities are markedly more frequent in HRCT patterns associated with advanced fibrosis.[4] Excluding B-lines, therefore, may delay diagnosis and hinder early screening efforts.

Moreover, a significant association has been observed between the number of B-lines and the nonspecific interstitial pneumonia (NSIP) pattern, which may help distinguish it from usual interstitial pneumonia.[4] The presence of B-lines has also been linked to inflammatory phases of the disease.[4] Ignoring B-lines could limit diagnostic accuracy—considering that NSIP is the predominant ILD pattern in SSc (affecting between 50% and 77.5% of cases)[5]—and negatively impact therapeutic decisions (eg, deciding between optimizing immunosuppression or initiating antifibrotic therapy) as well as prognostic assessment.

It would also have been interesting to evaluate the selected criteria according to radiologic pattern. Organizing pneumonia (OP), in general, presents with multiple—usually bilateral—subpleural and basally distributed pulmonary consolidations, ground-glass opacities, and other signs but typically lacks honeycombing.[2] There is limited evidence regarding the use of ultrasound for diagnosing OP. However, considering the aforementioned, relying solely on pleural irregularities or thickening—features associated with honeycombing and traction bronchiectasis[4]—may present important limitations in identifying this specific entity.

Finally, one of the key advantages of LUS over HRCT is its suitability for repeated assessments. This has been previously recognized, particularly in treatment monitoring and follow-up, with the potential to reduce the number of HRCT scans needed.[4,6] Therefore, we believe that LUS should not be viewed as a competing technique to HRCT but rather as a complementary tool[7] to reduce unnecessary exposure to ionizing radiation.

We would like to thank the authors for their valuable work. Their findings once again underscore the utility of point-of-care ultrasound in monitoring patients with autoimmune diseases, enabling better clinical assessment in cases of deterioration and helping identify those at higher risk of ILD, thereby facilitating earlier diagnostic and therapeutic interventions. Continued refinement and validation of ultrasound-based criteria—integrating both pleural and parenchymal findings—will be essential to maximize its diagnostic accuracy and consolidate its role as a complementary tool to high-resolution imaging in routine clinical practice.

The authors acknowledge the minimal use of a large language model (LLM), specifically ChatGPT (OpenAI), for support in editing this manuscript. The LLM was used exclusively to check grammar, improve clarity, and ensure fluency of the English text. No content was generated or interpreted by the LLM, and all scientific concepts, data analyses, and interpretations were developed independently by the authors.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25606.

Pablo Echevarría Díez-Canedo, MD
Hospital Universitario Vall d'Hebron
Barcelona, Spain
Raquel Marín-Baselga, MD
Hospital Universitario La Paz
Madrid, Spain
Yale Tung-Chen, MD, PhD
yale.tung@uam.es
Hospital Universitario La Paz
and Universidad Autónoma de Madrid
Madrid, Spain

1. Fairchild RM, Mar DA, Deluna MD, et al. Validation of lung ultrasound interpretation criteria for interstitial lung disease in systemic sclerosis and inflammatory myopathy. Arthritis Care Res (Hoboken) 2025;77(11):1384–1392.

2. Biciusca V, Rosu A, Stan SI, et al. A practical multidisciplinary approach to identifying interstitial lung disease in systemic autoimmune rheumatic diseases: a clinician's narrative review. Diagnostics (Basel) 2024;14(23):2674.
3. Bruni C, Mattolini L, Tofani L, et al. Lung ultrasound B-lines in the evaluation of the extent of interstitial lung disease in systemic sclerosis. Diagnostics (Basel) 2022;12(7):1696.
4. Rotondo C, Busto G, Rella V, et al. Transthoracic lung ultrasound in systemic sclerosis-associated interstitial lung disease: capacity to differentiate chest computed-tomographic characteristic patterns. Diagnostics (Basel) 2025;15(4):488.
5. Bastos AL, Corrêa RA, Ferreira GA. Tomography patterns of lung disease in systemic sclerosis. Radiol Bras 2016;49(5):316–321.
6. Mongodi S, Colombo A, Orlando A, et al. Combined ultrasound-CT approach to monitor acute exacerbation of interstitial lung disease. Ultrasound J 2020;12(1):27.
7. Sporea I. Ultrasound and computer tomography in abdominal emergencies - competitors or complementary tools? Med Ultrason 2025; 27(1):5–6.

DOI 10.1002/acr.25607

Reply

To the Editor:

We appreciate the thoughtful comments regarding our recent study on lung ultrasound (LUS) interpretation criteria for interstitial lung disease (ILD) in systemic sclerosis (SSc) and inflammatory myopathies (IM).[1] The utility of B-lines in detecting interstitial syndromes is well established, with decades of research supporting their relevance. However, we believe it is important to contextualize these findings within the landscape of evolving ultrasound technology and clinical practice.

Over the past two decades, there have been substantial advancements in ultrasound equipment. Contemporary rheumatology ultrasound examinations are typically performed using high-frequency linear probes (12–20 MHz) on machines equipped with advanced image-processing capabilities.[2] These modern systems differ significantly from older devices employed in early studies on ILD,[3–6] which commonly used lower-frequency (2–5 MHz) curvilinear probes and lacked sophisticated postprocessing algorithms. Although such setups were well suited for generating conspicuous B-lines, they lacked the resolution to assess pleural surface detail—a key target in early ILD detection.

Our study aimed to define and validate a practical set of interpretation criteria for current clinical practice, specifically tailored to the routine outpatient setting in which patients with SSc and IM are regularly screened for ILD during their lengthy "at-risk" period.[7,8] Rheumatologists, who maintain longitudinal relationships with these patients—often extending a decade or more—are ideally positioned to perform this screening using their standard ultrasound equipment. With typical rheumatology setups, pleural abnormalities can be directly visualized, while B-lines, which are particularly sensitive to system configuration,[9–11] may be inconsistent or absent. Typical rheumatology ultrasound setups using a linear probe, with moderate to high frequency,[2] and settings such as compound spatial imaging and proprietary postprocessing algorithms—designed in many cases to reduce artifacts—can significantly diminish the appearance of B-lines.[9–11] This setup contrasts with the curvilinear or phased-array probes and handheld devices used in the more acute setting for cardiology,[12] pulmonology,[13] or emergency medicine,[14] where identification of B-lines is critical in the assessment of congestive heart failure or inflammatory pulmonary conditions.

In our experience, reconfiguring machines to mimic these setups—by disabling compound imaging, reducing frequency, and switching to a curvilinear probe—can substantially increase the visibility of B-lines. However, doing so requires significant deviation from standard rheumatology practice[2] and sacrifices pleural detail. In contrast, we found that pleural irregularities were consistently detectable in patients with ILD, even when B-lines were absent.

We recognize the value of B-lines in certain clinical scenarios—such as acute care settings and in evaluating volume status or inflammatory interstitial processes. However, for routine outpatient screening of connective tissue disease–associated ILD using standard rheumatology ultrasound equipment, we believe that focusing on pleural features rather than artifacts offers a more robust, reproducible, and practical approach. Our interpretation criteria reflect this perspective and, we believe, provide a clinically useful framework for identifying ILD in this context without reliance on B-lines.

Lastly, with respect to our subset of patients with organizing pneumonia (OP). Of the eight patients with OP, seven were accurately identified on LUS by all readers, with the eighth patient having a normal ultrasound as interpreted by all readers. The one patient with false-negative OP had very mild computed tomographic imaging without peripheral involvement precluding LUS detection; B-lines were also not observed for this patient.

We appreciate the continued dialogue on this evolving field and remain committed to refining ultrasound-based approaches to ILD detection and monitoring.

A large language model was used solely for spell checking and grammar in this manuscript..

Robert M. Fairchild, MD, PhD, RhMSUS
rfairchi@stanford.edu
Stanford University School of Medicine
Stanford, CA
Lorinda Chung, MD, MS
Stanford University School of Medicine
Stanford, CA
and VA Palo Alto Health Care System
Palo Alto, CA

1. Fairchild RM, Mar DA, Deluna MD, et al. Validation of lung ultrasound interpretation criteria for interstitial lung disease in systemic sclerosis and inflammatory myopathy. Arthritis Care Res (Hoboken) 2025; 77(11):1384–1392.
2. Naredo E, D'Agostino MA, Conaghan PG, et al. Current state of musculoskeletal ultrasound training and implementation in Europe: results

of a survey of experts and scientific societies. Rheumatology (Oxford) 2010;49(12):2438–2443.
3. Gargani L, Bruni C, Romei C, et al. Prognostic value of lung ultrasound B-lines in systemic sclerosis. Chest 2020;158(4):1515–1525.
4. Hassan RI, Lubertino LI, Barth MA, et al. Lung ultrasound as a screening method for interstitial lung disease in patients with systemic sclerosis. J Clin Rheumatol 2019;25(7):304–307.
5. Moazedi-Fuerst FC, Kielhauser S, Brickmann K, et al. Sonographic assessment of interstitial lung disease in patients with rheumatoid arthritis, systemic sclerosis and systemic lupus erythematosus. Clin Exp Rheumatol 2015;33(4 suppl 91):S87–S91.
6. Barskova T, Gargani L, Guiducci S, et al. Lung ultrasound for the screening of interstitial lung disease in very early systemic sclerosis. Ann Rheum Dis 2013;72(3):390–395.
7. Nihtyanova SI, Denton CP. Pathogenesis of systemic sclerosis associated interstitial lung disease. J Scleroderma Relat Disord 2020;5(2 suppl):6–16.
8. Hu Q, Huang KC, Goh CH, et al. Characteristics and risk of interstitial lung disease in dermatomyositis and polymyositis: a retrospective cohort study in Japan. Sci Rep 2023;13(1):17172.
9. Demi M, Prediletto R, Soldati G, et al. Physical mechanisms providing clinical information from ultrasound lung images: hypotheses and early confirmations. IEEE Trans Ultrason Ferroelectr Freq Control 2020; 67(3):612–623.
10. Dietrich CF, Mathis G, Blaivas M, et al. Lung B-line artefacts and their use. J Thorac Dis 2016;8(6):1356–1365.
11. Soldati G, Demi M, Inchingolo R, et al. On the physical basis of pulmonary sonographic interstitial syndrome. J Ultrasound Med 2016; 35(10):2075–2086.
12. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2015;16(3):233–270.
13. Volpicelli G, Elbarbary M, Blaivas M, et al; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med 2012;38(4):577–591.
14. Katz JF, Yucel EK. Point-of-care ultrasonography. N Engl J Med 2011;364(21):2075–2076.

DOI 10.1002/acr.25605

Beyond the bloodstream—rethinking colchicine monitoring through a clinical lens: comment on the article by Stamp et al

To the Editor:

We read with great interest the article recently published by Stamp et al, "Colchicine concentrations and relationship with colchicine efficacy and adverse events: post hoc analysis of a randomized clinical trial of colchicine for gout flare prophylaxis,"[1] which examined the relationship between plasma colchicine concentrations and both clinical efficacy and adverse events. Although the study concludes that trough and peak concentrations bear no relationship to flare prevention or colchicine-specific toxicity and thus that routine therapeutic drug monitoring (TDM) is of limited value, from a clinical perspective, I propose three considerations to refine how these findings are applied in practice.

First, relying solely on plasma levels may underestimate colchicine's pharmacodynamic activity because its anti-inflammatory effects are mediated intracellularly. Colchicine partitions extensively into leukocytes, binding to tubulin and disrupting neutrophil function at tissue sites, yet plasma concentrations represent only a small—and highly variable—fraction of total drug exposure.[2–4] In clinical practice, patients with identical plasma levels may experience divergent responses based on differences in cellular uptake, transporter polymorphisms, or tissue distribution. Future studies integrating measures of colchicine in leukocytes or inflammatory biomarkers (eg, interleukin-1β inhibition) could better capture the concentration–effect relationship and identify patients who might benefit from dose adjustments despite "therapeutic" plasma levels.

Second, the association between higher plasma concentrations and nonspecific adverse events highlights the interplay of patient factors beyond colchicine exposure itself. In this analysis, elevated trough or peak levels correlate with any adverse event in months four to six, yet gastrointestinal and muscle-specific toxicity remain unlinked to plasma concentration. This suggests that comorbidities such as renal impairment, polypharmacy (statin coadministration), and older age—each associated with altered colchicine clearance and vulnerability to side effects—may drive overall adverse event reporting rather than colchicine per se. Clinicians should therefore interpret plasma concentrations in the context of individual risk factors; for example, in an older patient with chronic kidney disease and multiple cytochrome P450 3A4 or P-glycoprotein inhibitors, modest elevations in plasma colchicine may portend general toxicity risk even if specific adverse events are not evident.

Finally, translating these pharmacokinetic observations into clinical practice would benefit from alignment with patient-centered outcomes and shared decision-making. The absence of a plasma concentration threshold predictive of flare prevention underscores that gout prophylaxis success ultimately rests on clinical metrics (flare frequency and patient-acceptable disease activity states) and tolerability. Rather than routine TDM, a strategy of early assessment of clinical response and targeted inquiry about gastrointestinal or neuromyotoxic symptoms may be more effective. Patients could be counseled that dose titration should be guided by flare control and symptom burden, with plasma measurements reserved for those with suspected toxicity or drug–drug interactions. Such an approach prioritizes real-world outcomes and resource stewardship, consistent with the finding that concentration monitoring added little value across an unselected population.

In summary, although the study robustly demonstrates that steady-state plasma colchicine concentrations do not predict prophylactic efficacy or specific toxicity on a population level, incorporating insights into intracellular drug action, individual comorbidity profiles, and patient-reported outcomes can optimize the clinical

of a survey of experts and scientific societies. Rheumatology (Oxford) 2010;49(12):2438–2443.
3. Gargani L, Bruni C, Romei C, et al. Prognostic value of lung ultrasound B-lines in systemic sclerosis. Chest 2020;158(4):1515–1525.
4. Hassan RI, Lubertino LI, Barth MA, et al. Lung ultrasound as a screening method for interstitial lung disease in patients with systemic sclerosis. J Clin Rheumatol 2019;25(7):304–307.
5. Moazedi-Fuerst FC, Kielhauser S, Brickmann K, et al. Sonographic assessment of interstitial lung disease in patients with rheumatoid arthritis, systemic sclerosis and systemic lupus erythematosus. Clin Exp Rheumatol 2015;33(4 suppl 91):S87–S91.
6. Barskova T, Gargani L, Guiducci S, et al. Lung ultrasound for the screening of interstitial lung disease in very early systemic sclerosis. Ann Rheum Dis 2013;72(3):390–395.
7. Nihtyanova SI, Denton CP. Pathogenesis of systemic sclerosis associated interstitial lung disease. J Scleroderma Relat Disord 2020;5(2 suppl):6–16.
8. Hu Q, Huang KC, Goh CH, et al. Characteristics and risk of interstitial lung disease in dermatomyositis and polymyositis: a retrospective cohort study in Japan. Sci Rep 2023;13(1):17172.
9. Demi M, Prediletto R, Soldati G, et al. Physical mechanisms providing clinical information from ultrasound lung images: hypotheses and early confirmations. IEEE Trans Ultrason Ferroelectr Freq Control 2020; 67(3):612–623.
10. Dietrich CF, Mathis G, Blaivas M, et al. Lung B-line artefacts and their use. J Thorac Dis 2016;8(6):1356–1365.
11. Soldati G, Demi M, Inchingolo R, et al. On the physical basis of pulmonary sonographic interstitial syndrome. J Ultrasound Med 2016; 35(10):2075–2086.
12. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2015;16(3):233–270.
13. Volpicelli G, Elbarbary M, Blaivas M, et al; International Liaison Committee on Lung Ultrasound (ILC-LUS) for International Consensus Conference on Lung Ultrasound (ICC-LUS). International evidence-based recommendations for point-of-care lung ultrasound. Intensive Care Med 2012;38(4):577–591.
14. Katz JF, Yucel EK. Point-of-care ultrasonography. N Engl J Med 2011;364(21):2075–2076.

DOI 10.1002/acr.25605

Beyond the bloodstream—rethinking colchicine monitoring through a clinical lens: comment on the article by Stamp et al

To the Editor:

We read with great interest the article recently published by Stamp et al, "Colchicine concentrations and relationship with colchicine efficacy and adverse events: post hoc analysis of a randomized clinical trial of colchicine for gout flare prophylaxis,"[1] which examined the relationship between plasma colchicine concentrations and both clinical efficacy and adverse events. Although the study concludes that trough and peak concentrations bear no relationship to flare prevention or colchicine-specific toxicity and thus that routine therapeutic drug monitoring (TDM) is of limited value, from a clinical perspective, I propose three considerations to refine how these findings are applied in practice.

First, relying solely on plasma levels may underestimate colchicine's pharmacodynamic activity because its anti-inflammatory effects are mediated intracellularly. Colchicine partitions extensively into leukocytes, binding to tubulin and disrupting neutrophil function at tissue sites, yet plasma concentrations represent only a small—and highly variable—fraction of total drug exposure.[2–4] In clinical practice, patients with identical plasma levels may experience divergent responses based on differences in cellular uptake, transporter polymorphisms, or tissue distribution. Future studies integrating measures of colchicine in leukocytes or inflammatory biomarkers (eg, interleukin-1β inhibition) could better capture the concentration–effect relationship and identify patients who might benefit from dose adjustments despite "therapeutic" plasma levels.

Second, the association between higher plasma concentrations and nonspecific adverse events highlights the interplay of patient factors beyond colchicine exposure itself. In this analysis, elevated trough or peak levels correlate with any adverse event in months four to six, yet gastrointestinal and muscle-specific toxicity remain unlinked to plasma concentration. This suggests that comorbidities such as renal impairment, polypharmacy (statin coadministration), and older age—each associated with altered colchicine clearance and vulnerability to side effects—may drive overall adverse event reporting rather than colchicine per se. Clinicians should therefore interpret plasma concentrations in the context of individual risk factors; for example, in an older patient with chronic kidney disease and multiple cytochrome P450 3A4 or P-glycoprotein inhibitors, modest elevations in plasma colchicine may portend general toxicity risk even if specific adverse events are not evident.

Finally, translating these pharmacokinetic observations into clinical practice would benefit from alignment with patient-centered outcomes and shared decision-making. The absence of a plasma concentration threshold predictive of flare prevention underscores that gout prophylaxis success ultimately rests on clinical metrics (flare frequency and patient-acceptable disease activity states) and tolerability. Rather than routine TDM, a strategy of early assessment of clinical response and targeted inquiry about gastrointestinal or neuromyotoxic symptoms may be more effective. Patients could be counseled that dose titration should be guided by flare control and symptom burden, with plasma measurements reserved for those with suspected toxicity or drug–drug interactions. Such an approach prioritizes real-world outcomes and resource stewardship, consistent with the finding that concentration monitoring added little value across an unselected population.

In summary, although the study robustly demonstrates that steady-state plasma colchicine concentrations do not predict prophylactic efficacy or specific toxicity on a population level, incorporating insights into intracellular drug action, individual comorbidity profiles, and patient-reported outcomes can optimize the clinical

use of colchicine. This nuanced perspective will ensure that therapy remains both effective and safe for those at the highest risk.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25605.

Yadi Li, MD
yadili9232@gmail.com
Zheng Wei, MD
Department of Traditional Chinese Medicine
Deyang People's Hospital
Deyang, China
Jianlong Zhou, MD
Chengdu University of Traditional Chinese Medicine
Chengdu, China

1. Stamp LK, Horne A, Mihov B, et al. Colchicine concentrations and relationship with colchicine efficacy and adverse events: post hoc analysis of a randomized clinical trial of colchicine for gout flare prophylaxis. Arthritis Care Res 2025;77(9):1157–1164.
2. Niel E, Scherrmann JM. Colchicine today. Joint Bone Spine 2006; 73(6):672–678.
3. Forkosh E, Kenig A, Ilan Y. Introducing variability in targeting the microtubules: review of current mechanisms and future directions in colchicine therapy. Pharmacol Res Perspect 2020;8(4):e00616.
4. Paschke S, Weidner AF, Paust T, et al. Technical advance: inhibition of neutrophil chemotaxis by colchicine is modulated through viscoelastic properties of subcellular compartments. J Leukoc Biol 2013;94(5): 1091–1096.