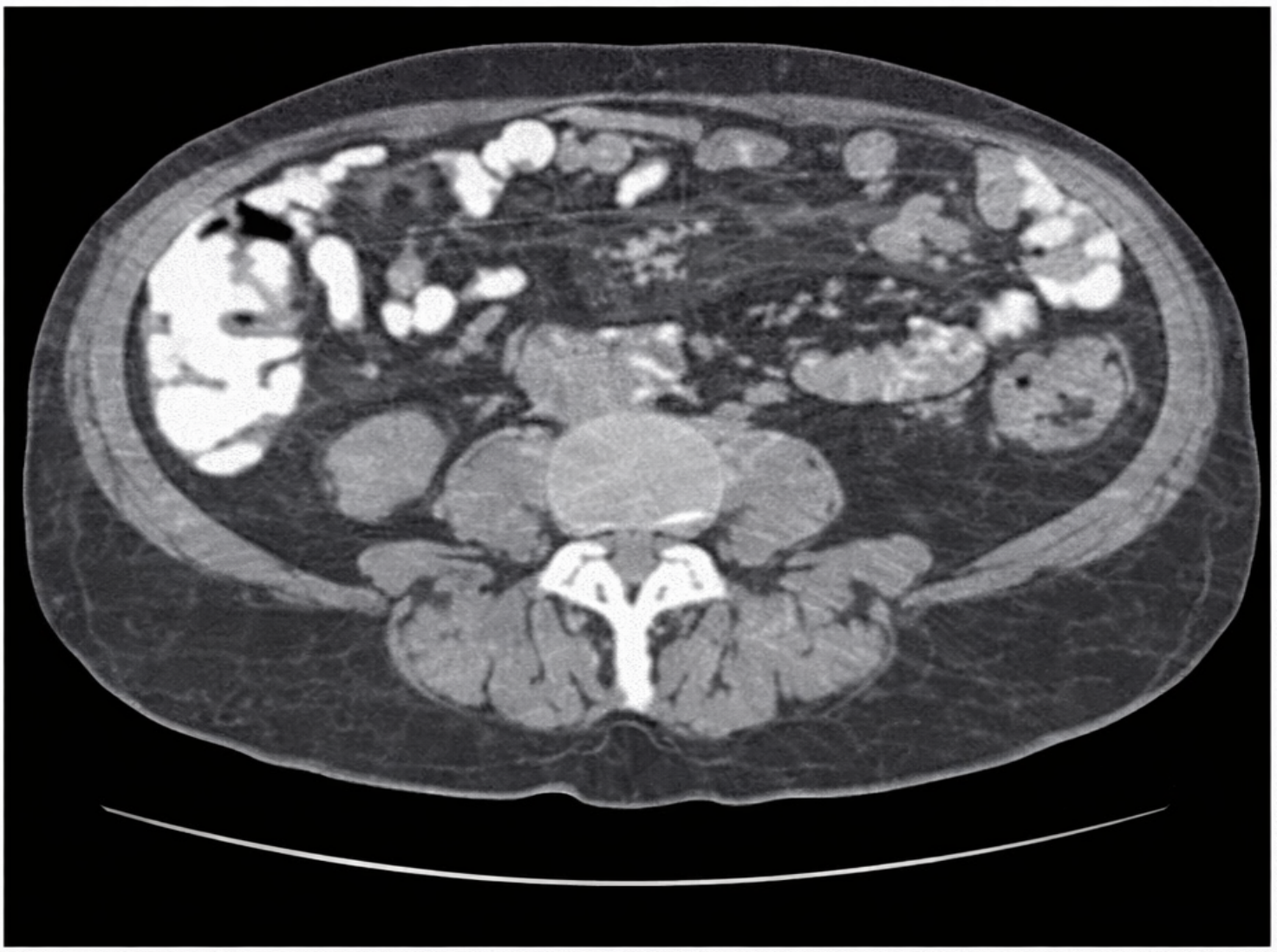
American College of Rheumatology

WILEY

Arthritis Care & Research

AN OFFICIAL JOURNAL OF THE
AMERICAN COLLEGE OF RHEUMATOLOGY

VOLUME 72 • NUMBER 2
FEBRUARY 2020



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. 2 February 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 • FEBRUARY 2026 • NO. 2

Cover image: The image on the cover (Sondhi et al; pages 182–187) shows a CT of the abdomen/pelvis with contrast showing a periaortic lesion abutting the inferior vena cava, consistent with a large retroperitoneal lymph node, along with retroperitoneal mesenteric lymphadenopathy and fat stranding.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 177–181
DOI 10.1002/acr.25638

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Uses of Artificial Intelligence in Nonpharmacological Rheumatology Care: What We Know, What We Need to Learn, and What Its Limitations Are

Travis Haber, Rana S. Hinman, and Mark Merolli

Artificial intelligence (AI) broadly refers to intelligent systems—whether human-like or rational—that can process information from the environment to make decisions and work autonomously toward defined goals.[1] Rules-based AIs operate on predefined rules (eg, if/then logic), designed to make decisions like human experts. These AIs are among the earliest adopted in health care. In contrast, machine learning, and its more advanced subtype, deep learning, can perform classification and predictive modeling of data, without requiring researchers to specify rules in advance.[2] Building on these capabilities, research is now focusing on the clinical applications of these AI methods, notably for analyzing health imaging to diagnose health conditions (such as magnetic resonance imaging to diagnose rheumatoid arthritis) and predicting a patient's prognosis and treatment response by modeling laboratory and electronic medical record data (eg, risk of progression to arthroplasty).[2–4] More recently, deep learning has been used to develop large language models (LLMs), which are being increasingly evaluated to assess whether they can effectively diagnose rheumatologic conditions and provide treatment recommendations aligned with clinical guidelines (eg, ChatGPT by OpenAI).[2,5] These rapidly advancing AI models may have the potential to deliver health advice and support rheumatology care at scale.[6] Taken together, integrating AI into interventions could usher in an era of truly personalized "precision" medicine—in which AI integrates and analyses multiple sources of data (eg, clinical laboratory and omics data) to predict which treatments are most likely to benefit a given patient.[3,4]

In light of this rapidly changing field, Shah et al conducted a scoping review of AI-supported interventions for the nonpharmacological management of rheumatic diseases, featured in this issue of *Arthritis Care & Research*.[7] This review of 15 studies found that most were focused on osteoarthritis (OA) care (n = 11, 73%), with most aiming to educate participants about the condition and provide guidance on exercise management (n = 10, 67%). Rules-based systems were the most common type of AI integrated into interventions. Most studies were composed of pilot/feasibility or developmental/protocol papers (n = 10, 67%). Few fully powered randomized controlled trials (RCTs) were identified (n = 4, 27%), and, excluding abstracts, none evaluated the effects of interventions supported by deep learning (including LLMs) on clinical outcomes. Overall, this timely review highlights a growing but still nascent evidence base for the use of AI in nonpharmacological care for rheumatologic conditions. Expanding on these results, we highlight (1) AI uses in nonpharmacological rheumatology care that show early promise, (2) key limitations and challenges of AI in nonpharmacological rheumatology care, and (3) a future outlook for integrating AI into rheumatology care. Figure 1 summarizes these key themes. Ethics approval was not required for this editorial.

AI uses in nonpharmacological rheumatology care that show early promise

Rules-based digital systems (decision support). Interventions using rules-based systems are currently the only type of AI evaluated in peer-reviewed RCTs for their effects on clinical outcomes. Findings from three RCTs and one pilot RCT provide preliminary evidence that digital interventions using AI to personalize nonpharmacological care may improve physical activity and some outcomes, such as pain and physical function.[8–10] Although these early findings are promising, the effects of these interventions on clinical outcomes have been inconsistent across trials. Furthermore, only one study has assessed the long-term impact of an AI-supported intervention, finding no difference in pain, function, or quality of life.[10] Further RCTs are therefore needed to determine

Dr Hinman's work was supported by a National Health and Medical Research Council Investigator Grant (2025733).

Travis Haber, PhD, Rana S. Hinman, PhD, Mark Merolli, PhD: Centre for Health Exercise, and Sports Medicine, Department of Physiotherapy, School of Health Sciences, The University of Melbourne, Melbourne, Australia.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25638.

Address correspondence via email to Travis Haber, PhD, at travis.haber@unimelb.edu.au.

Submitted for publication August 8, 2025; accepted in revised form August 8, 2025.

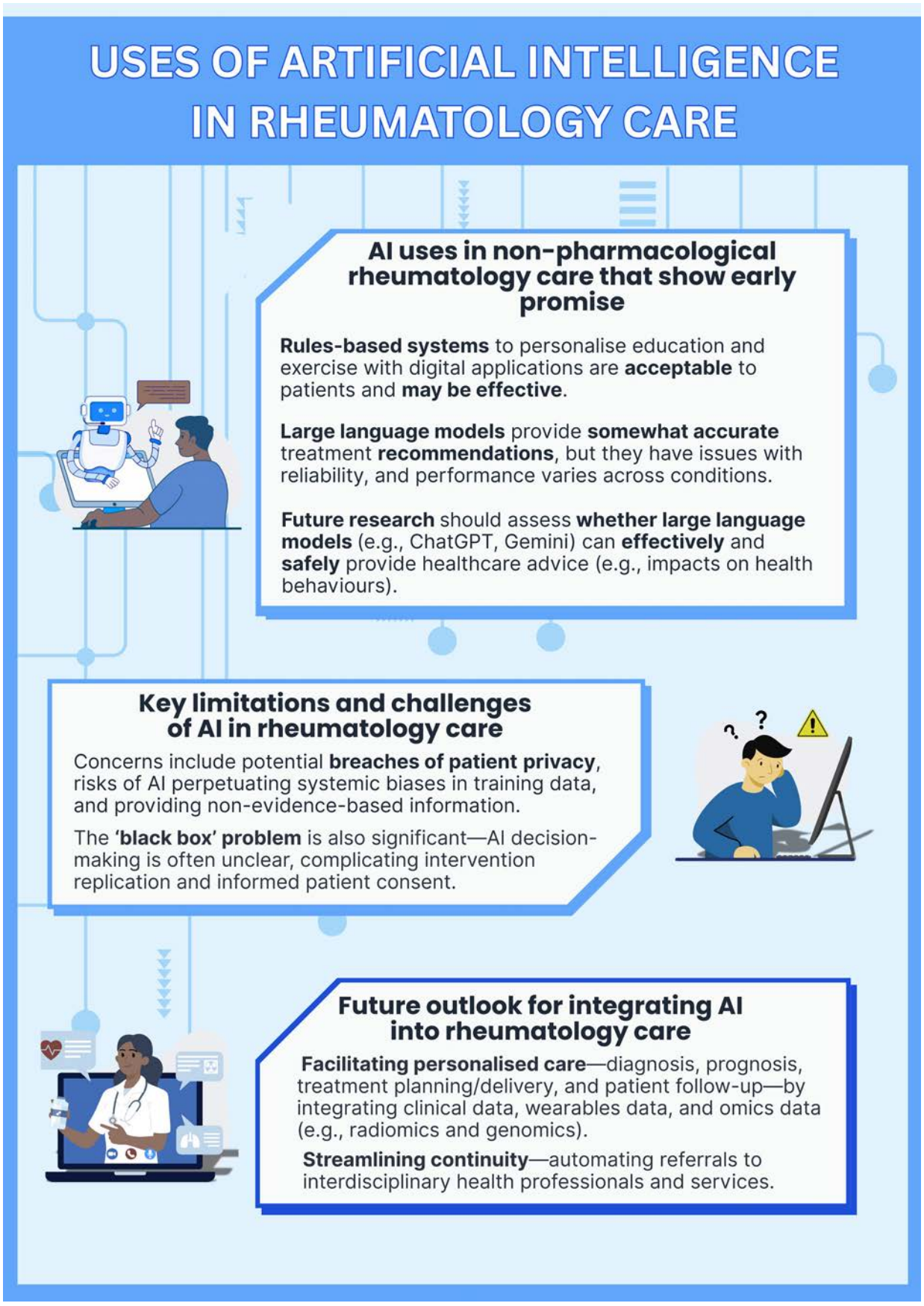


Figure 1. Uses of artificial intelligence in rheumatology care. AI, artificial intelligence.

whether rules-based AI-supported interventions meaningfully improve clinical outcomes and whether these effects are sustained long term. With the increasing research focus on digital interventions for rheumatology care, such as in OA,[11] integrating rules-based AI into such interventions may help transform these strategies from a "one-size-fits-all" approach to more tailored care for patients. Research is also needed to determine how these interventions can be implemented at scale to diverse patient populations with varying digital literacy and preferences for digital care.[12]

Chatbots and conversational agents. LLMs, such as ChatGPT, can enable complex, interactive communication between users and systems, sometimes through agent-based interfaces.[13] Two studies have evaluated whether LLMs can reliably and accurately recommend rheumatologic care to patients (primarily for OA).[14,15] The reliability and usefulness of information provided by ChatGPT (version unspecified) varied by condition—for example, it was most reliable for OA and most useful for ankylosing spondylitis.[15] Furthermore, ChatGPT (version not reported) and Bard (now Gemini [Google]) differed in their concordance with recommendations in clinical guidelines.[14] These findings are similar to a study in low back pain that also found that ChatGPT 3.5, Bing, Bard (Gemini), and ChatGPT 4.0 performed modestly in terms of accuracy (about 56% accurate) and had limitations with reliability.[16] Collectively, emerging evidence suggests that publicly available LLMs can provide somewhat accurate treatment recommendations for rheumatologic conditions. However, further work is needed to improve the accuracy and reliability of LLMs in delivering rheumatologic treatment advice.

The general public is increasingly using AI to seek health information through engagement with LLMs.[17] A survey in the United States and Puerto Rico (n = 2,406) found that approximately 20% of participants used ChatGPT to seek health information, and most used the information to guide their health behaviors (such as seeking a referral or changing medications).[17] Future RCTs are needed to evaluate whether these LLMs provide accurate and reliable information, are safe, and positively influence health behaviors.[18] Consumer preference should also guide future research. Patients appear to prefer human-like chatbots (eg, digital avatars capable of simulating empathy) over machine-like conversational agents.[18] Lastly, future research should evaluate how best to integrate chatbots and conversational agents into clinical care, which could include hybrid interventions (combining them with clinician-delivered care) or stepped interventions (in which the AI intervention is delivered first, with an intervention from a clinician only provided for participants who need it).

Key limitations and challenges of AI in nonpharmacological rheumatology care

Although AI is likely to shape health care in the coming decade, it is not without important limitations and challenges. Key concerns surrounding AI include the potential to breach patient privacy, as well as the risk that AI may perpetuate systemic biases contained in training datasets against traditionally marginalized groups of people.[4,19,20] Increasing reliance on AI for health care could therefore widen existing health disparities, particularly if efforts are not made to ensure equitable access to AI, as well as to train AI models that can provide culturally appropriate care.[21] Other concerns relate to the accuracy of the information that AI provides. For example, LLMs are typically trained on data from the internet, such as blogs, news articles, websites, books, scientific articles, and guidelines, which makes them susceptible to providing nonevidence-based recommendations for care.[16] These models are also prone to confabulate (hallucinate) treatment advice (ie, recommendations not contained within the training set).[14,22]

Another key concern relating to AI is the so-called "black-box" problem. Broadly, this problem refers to the fact that it is not always clear how AI makes a recommendation (eg, algorithms, inputs, and outputs are not always known to the user).[4,22,23] This means that AI can be opaque and difficult to replicate across different populations and contexts. Although numerous guidelines, such as the Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence (DECIDE-AI) guidelines,[24] have been developed to improve the quality of reporting of AI studies, the use of such guidelines remains infrequent. Opaqueness may also mean that a researcher or clinician cannot describe to a patient how the clinical decision was arrived at.[23] Can that patient, then, truly provide informed consent? Relatedly, the degree to which patients can opt in and out of AI is relatively straightforward in the research setting, but it may be less so in a situation in which an AI-supported intervention is scaled and embedded in routine health care decision-making.[23] Another challenge is the rapid advancement of AI. For example, ChatGPT (OpenAI) releases a new model annually or more frequently, which potentially outpaces the ability of researchers to rigorously evaluate the most current models available to the public. As such, innovative research approaches will be required to keep pace with advances in AI over the coming years.

A future outlook for integrating AI into rheumatology care

There are various ways in which AI may advance health care in the future. AI could help streamline continuity of care by automating referrals to health professionals and services, along with performing risk stratification and patient support between clinician visits.[6,25–27] It could also automate workflows and health care services to alleviate the strain faced by global, resource-constrained health systems, for example, by automating clinical notes and prescribing medications.[25] AI is poised to enable personalized care (potentially around the clock) that is often lacking in health care. In cancer and cardiovascular diseases, AI systems

are already being developed to integrate information from clinicians' assessments, wearable devices (eg, sleep, metabolic markers, physical activity), and omics data, such as radiomics, proteomics, genomics, and metabolomics.[26,28] Such systems, if adapted to rheumatology care, could assist in diagnosing conditions (eg, analyzing medical imaging), predicting prognosis (eg, risk of progression to arthroplasty), planning and recommending treatments, performing follow-ups, and supporting long-term adherence to management.[3,4,26,28] Incredibly, AI may soon enable the testing of a treatment on a patient's hypothetical digital twin to see how they might respond to a given treatment.[2,26] Importantly, if developed and implemented ethically and thoughtfully, AI could help democratize person-centered care at scale, particularly for patients who traditionally experience barriers to in-person, supervised care, such as those with limited financial resources or who live remotely.[21]

Conclusions

In summary, as highlighted by the scoping review by Shah et al (2025),[7] research evaluating the effectiveness of AI-supported interventions has mostly focused on rules-based systems to personalize digital care. Although early evidence suggests that these AI applications may improve some outcomes, future research is needed to explore the broader applications of AI in rheumatologic care, such as whether AI can effectively support positive behavioral changes and provide personalized patient 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 Haber 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. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach. Global Edition 4e. Pearson Education; 2021.
2. Sequí-Sabater JM, Benavent D. Artificial intelligence in rheumatology research: what is it good for? RMD Open 2025;11(1):e004309.
3. Sun Y, Lin J, Chen W. Artificial intelligence in rheumatoid arthritis. Rheumatol Autoimmun 2025;5(2):88–100.
4. Ou J, Zhang J, Alswadeh M, et al. Advancing osteoarthritis research: the role of AI in clinical, imaging and omics fields. Bone Res 2025; 13(1):48.
5. Lawford BJ, Bennell KL, Haber T, et al. Osteoarthritis Year In Review 2024: rehabilitation and outcomes. Osteoarthritis Cartilage 2024; 32(11):1405–1412.
6. Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine. Commun Med (Lond) 2023; 3(1):141.
7. Shah N, Castellanos A, Chen Y, et al. A scoping review on ai-supported interventions for non-pharmacological management of chronic rheumatic diseases. Arthritis Care Res (Hoboken) 2026; 78(2):270–281.
8. Rini C, Porter LS, Somers TJ, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015;156(5):837–848.
9. Pelle T, Bevers K, van der Palen J, et al. Effect of the dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthritis Cartilage 2020;28(4):418–427.
10. Bossen D, Veenhof C, Van Beek KE, et al. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013; 15(11):e257.
11. Lawford BJ, Hall M, Hinman RS, et al. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2024;12(12):CD004376.
12. Nelligan RK, Hinman RS, Teo PL, et al. Exploring attitudes and experiences of people with knee osteoarthritis toward a self-directed eHealth intervention to support exercise: qualitative study. JMIR Rehabil Assist Technol 2020;7(2):e18860.
13. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018;25(9): 1248–1258.
14. Yang J, Ardavanis KS, Slack KE, et al. Chat Generative Pretrained Transformer (ChatGPT) and Bard: artificial intelligence does not yet provide clinically supported answers for hip and knee osteoarthritis. J Arthroplasty 2024;39(5):1184–1190.
15. Uz C, Umay E. "Dr ChatGPT": is it a reliable and useful source for common rheumatic diseases? Int J Rheum Dis 2023;26(7):1343–1349.
16. Scaff SPS, Reis FJJ, Ferreira GE, et al. Assessing the performance of AI chatbots in answering patients' common questions about low back pain. Ann Rheum Dis 2025;84(1):143–149.
17. Ayo-Ajibola O, Davis RJ, Lin ME, et al. Characterizing the adoption and experiences of users of artificial intelligence-generated health information in the United States: cross-sectional questionnaire study. J Med Internet Res 2024;26:e55138.
18. Anisha SA, Sen A, Bain C. Evaluating the potential and pitfalls of AI-powered conversational agents as humanlike virtual health carers in the remote management of noncommunicable diseases: scoping review. J Med Internet Res 2024;26:e56114.
19. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019; 366(6464):447–453.
20. Kanter GP, Packel EA. Health care privacy risks of AI chatbots. JAMA 2023;330(4):311–312.
21. Laymouna M, Ma Y, Lessard D, et al. Roles, users, benefits, and limitations of chatbots in health care: rapid review. J Med Internet Res 2024;26:e56930.
22. De Cassai A, Dost B. Concerns regarding the uncritical use of ChatGPT: a critical analysis of AI-generated references in the context of regional anesthesia. Reg Anesth Pain Med 2024;49(5):378–380.
23. Xu H, Shuttleworth KMJ. Medical artificial intelligence and the black box problem: a view based on the ethical principle of "do no harm." Intell Med 2024;4(1):52–57.
24. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 2022;377:e070904.

25. Sarkar U, Bates DW. Using artificial intelligence to improve primary care for patients and clinicians. JAMA Intern Med 2024;184(4): 343–344.
26. Singh M, Kumar A, Khanna NN, et al. Artificial intelligence for cardiovascular disease risk assessment in personalised framework: a scoping review. EClinicalMedicine 2024;73:102660.
27. Merolli M, Gray K, Choo D, et al. Use, and acceptability, of digital health technologies in musculoskeletal physical therapy: a survey of physical therapists and patients. Musculoskeletal Care 2022;20(3):641–659.
28. Ghosh S, Zhao X, Alim M, et al. Artificial intelligence applied to 'omics data in liver disease: towards a personalised approach for diagnosis, prognosis and treatment. Gut 2025;74(2):295–311.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 182–187
DOI 10.1002/acr.25624

CLINICOPATHOLOGIC CONFERENCE

A Conundrum in Coexistence

Manush Sondhi and Samina Hayat

CASE PRESENTATION

Chief symptoms

The patient, a 52-year-old man, presented to his primary care physician with a history of persistent nasal congestion and rhinorrhea, which had been ongoing for several years but had recently worsened, significantly affecting his daily activities.

History of the present illness

The patient reported that his nasal congestion, along with a thick yellow-green nasal discharge, postnasal drip, and persistent cough, had gradually worsened over the past few months. He also experienced intermittent headaches localized to the frontal region. He denied having any fever, shortness of breath, wheezing, chest tightness, dental pain, ear problems, drooling, mouth sores, nosebleeds, or sore throat during this period.

Past medical history

His medical history includes hypertension, asthma, and coronary artery disease, for which he underwent coronary artery bypass grafting and is currently on ticagrelor therapy. He also has a history of ischemic stroke, which occurred a few years ago, likely secondary to atherosclerotic disease in the setting of longstanding hypertension and coronary artery disease. He has since made a full recovery.

Social and family history

His social and family history was unremarkable, with no significant exposures, habits, or familial predispositions.

Allergy history

The patient has a history of well-controlled asthma, which was diagnosed in adulthood and is currently managed with albuterol and inhaled glucocorticoids. Based on the history provided, the severity of asthma appears to be intermittent, as symptoms occur occasionally without persistent or consistent patterns that would suggest a more severe classification. He was diagnosed with a hypersensitivity to aspirin in childhood, which provoked chest tightness and shortness of breath upon exposure.

Review of systems

The patient denies weight loss, fatigue, fever, rash, skin changes, dryness, redness, vision changes, joint pain or swelling, sicca symptoms, parotid or submandibular swelling, chest pain, palpitations, cough, shortness of breath, abdominal pain, nausea, vomiting, diarrhea, muscle weakness, neuropathy, headache, dizziness, easy bruising, depression, or anxiety.

Physical examination

On examination, his vital signs were stable: his blood pressure was 120/80 mm Hg, his heart rate was 86 beats per minute, his respiratory rate was 17 breaths per minute, his temperature was 36.3°C, and his oxygen saturation was 98% on room air. His body mass index was calculated as 29.32.

The head, eyes, ears, nose, and throat examination revealed no visible deformities or trauma to the head. The conjunctivae were clear without scleral icterus, and tympanic membranes were clear bilaterally, showing no erythema or bulging. The nasal passages displayed visible congestion and probable nasal polyps, with erythematous and swollen mucosa but no purulent discharge. There was tenderness on palpation over the frontal and maxillary sinuses bilaterally. The pharynx appeared mildly erythematous without exudate, and the tonsils were not enlarged. Lacrimal, parotid, and submandibular gland examinations were

Manush Sondhi, MD, Samina Hayat, MD: Louisiana State University, Shreveport.
Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25624.

Address correspondence via email to Manush Sondhi, MD, at manushsondhi@gmail.com.
Submitted for publication November 11, 2024; accepted in revised form July 28, 2025.

performed, and results were normal, with no evidence of enlargement, tenderness, or nodularity.

Lung auscultation was unremarkable, with clear breath sounds and no wheezing, rhonchi, or rales. The cardiovascular examination showed a regular rhythm with normal S1 and S2 sounds and no murmurs, rubs, or gallops. His skin was warm and dry, without erythema or rash. The abdomen was nondistended, with normal bowel sounds and no tenderness. Peripheral pulses were strong and 2+ bilaterally, and he had a full range of motion in all extremities without deformities or limitations. Neurologically, he was alert and oriented, with intact cranial nerves, sensory perception, and coordination, and his mood, thought processes, judgment, and insight were appropriate.

Laboratory evaluation

The patient's laboratory results showed a hemoglobin level of 15.7 g/dL (reference 12–16 g/dL) and a hematocrit level of 48.1% (reference 36%–48%). The white blood cell count was 9,500/μL (3,900–12,700/μL), with a neutrophil percentage of 64.4% (reference 38%–73%), a lymphocyte percentage of 15.4%, a monocyte percentage of 11.3% (reference 4%–15%), and a basophil percentage of 0.6% (reference 0%–1.9%). Marked eosinophilia was noted, with an eosinophil count of 1,300/μL (reference 30–350/μL) and an absolute eosinophil count (AEC) of 12,400/μL (reference 0–500/μL). The platelet count was 210,000/μL (150,000–400,000/μL), and the complete metabolic panel was unremarkable.

The patient was referred to allergy and immunology due to an elevated eosinophil count. The workup revealed markedly elevated IgE levels at 1,470 IU/mL (reference <100 IU/mL). The result of an allergic bronchopulmonary aspergillosis (ABPA) panel was negative. Tryptase levels were measured to rule out mast cell disorders and were found to be elevated at 33.4 ng/mL (reference <11.5 ng/mL). Results of serologic tests for parasitic infections, including *Toxocara* and *Strongyloides*, along with a stool analysis and fungal-specific serologic tests, were all negative.

The patient's autoimmune and serologic panel was negative for antinuclear antibodies, perinuclear and cytoplasmic antineutrophil cytoplasmic antibodies (ANCAs), myeloperoxidase antibodies, and glomerular basement membrane antibodies. Complement levels were within normal limits (C3 level at 77 mg/dL, C4 level at 14 mg/dL), and celiac testing was negative for antigliadin antibodies (IgG level at 9 U and IgA level at 2 U, both within reference range). All relevant infections, including syphilis, helminthic infections, tuberculosis, and hepatitis C virus, were thoroughly evaluated and tested, with negative results.

CLINICAL COURSE

The patient was initially referred to otorhinolaryngology, where a nasal endoscopy revealed persistent nasal polyposis, classified as grade 4. A computed tomography (CT) scan of the sinuses demonstrated severe pansinusitis, with dense material noted in the right maxillary and sphenoid sinuses. The scan also showed near-complete opacification of the maxillary sinuses, ethmoid air cells, sphenoid, and frontal sinuses, consistent with chronic and extensive inflammatory sinus disease (Figure 1). A CT scan of the chest revealed diffuse patchy small airway disease and areas of atelectasis. Although complete functional endoscopic sinus surgery was initially recommended, the patient chose conservative management and was started on antihistamines, nasal saline rinses, and decongestants to alleviate symptoms.

The patient was referred to allergy and immunology for further evaluation. Spirometry revealed mild airway obstruction, consistent with the patient's history of asthma. Given the elevated IgE levels and findings on spirometry, allergen immunotherapy was suggested as a long-term management option; however, the patient declined. As a result, symptomatic treatment with antihistamines, topical decongestants, and nasal saline rinses was continued to address nasal and airway inflammation and improve respiratory symptoms.

The patient was also referred to hematology/oncology to investigate the eosinophilia. Flow cytometry of the peripheral

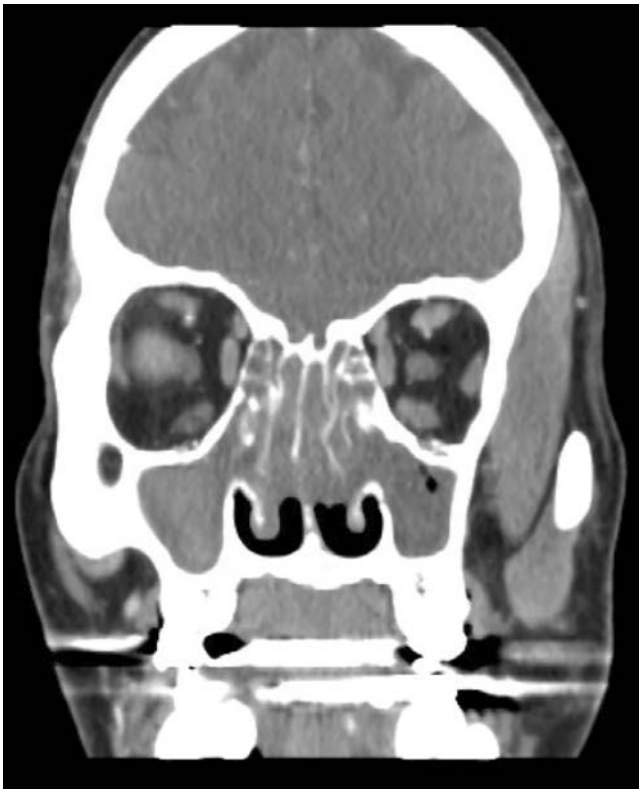


Figure 1. Sinus computed tomography scan revealing severe pansinusitis with dense material in the right maxillary and sphenoidal sinuses, along with severe mucosal disease and opacification of the maxillary sinuses, ethmoid cells, sphenoid sinus, and frontal sinus.

blood showed no markers of acute leukemia or non-Hodgkin lymphoma, and no evidence of clonal expansion was found, largely excluding malignant causes. A bone marrow biopsy confirmed a reactive, nonclonal process with 70% trilineage hematopoiesis, indicating normal bone marrow function without signs of dysplasia or infiltration. Cytogenetic and molecular testing, including the JAK2 V617F mutation (to exclude myeloproliferative neoplasms), FIP1L1-PDGFRA (to rule out rare causes of hypereosinophilic syndrome), PDGFRB fusion transcripts (to assess for chronic eosinophilic leukemia), and BCR-ABL (to rule out the eosinophilic variant of chronic myelogenous leukemia), all returned negative results. In the absence of malignancy, hydroxyurea therapy was initiated at 1,000 mg daily to manage the eosinophilia for presumed idiopathic hypereosinophilic syndrome (IHES).

After he was lost to follow-up for almost two years, the patient returned to his primary care physician with new symptoms, including intermittent dull lower abdominal pain radiating to the lower back, which started a few months prior and worsened over a few days, along with intermittent diarrhea. There were no episodes of melena or hematochezia. The patient also reported worsening sinus-related symptoms that were not relieved by medications. A CT scan of the abdomen and pelvis with contrast revealed a periaortic lesion abutting the inferior vena cava, consistent with a large retroperitoneal lymph node, as well as retroperitoneal mesenteric lymphadenopathy with fat stranding (Figure 2). These findings prompted a referral to rheumatology for further evaluation.

Given the progression of sinus-related symptoms, the patient was referred to an ear, nose, and throat (ENT) specialist again, and this time, he ultimately agreed to undergo sinus surgery. Biopsy specimens taken from the polyps confirmed acute and chronic lymphoplasmacytic inflammation with storiform fibrosis, along with the presence of eosinophils and more than 10 IgG4 cells per high-power field, with an $IgG4^+/IgG^+$ cell ratio >40%.

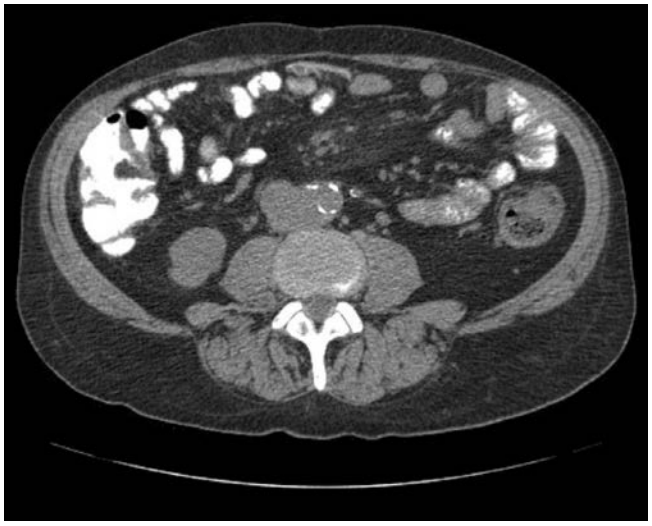


Figure 2. Computed tomography scan of the abdomen/pelvis with contrast showing a periaortic lesion abutting the IVC, consistent with a large retroperitoneal lymph node, along with retroperitoneal mesenteric lymphadenopathy and fat stranding.

Subsequent Ig panel results showed a markedly elevated total IgG level at 4,180 mg/dL (650–1,600 mg/dL), whereas IgA and IgM levels were within normal limits. An IgG subclass analysis revealed normal levels of IgG1, IgG2, and IgG3 and a markedly elevated IgG4 level at 2,413 mg/dL (4–86 mg/dL).

Following this diagnosis, the patient was started on a short course of prednisone, beginning at 40 mg, which was tapered rapidly. Two 1-g intravenous rituximab infusions, separated by two weeks, were initiated alongside it, with a plan to repeat the course every six months. Hydroxyurea was discontinued. The patient reported improvement in symptoms, including abdominal pain and sinus symptoms, following rituximab infusions. CT findings showed a reduction in retroperitoneal lymphadenopathy, and IgG4 levels decreased, indicating a positive response to therapy.

CASE SUMMARY

The patient, a 52-year-old man with a history of asthma, aspirin hypersensitivity, coronary artery disease, and ischemic strokes, presented with chronic nasal congestion, rhinorrhea, and cough for several years. The initial workup revealed eosinophilia with an eosinophil count of 1,300/μL and an AEC of 12,400/μL, along with elevated IgE levels. The patient was referred to ENT, where endoscopy showed grade 4 nasal polyposis, and CT demonstrated severe pansinusitis. Despite recommendations for complete functional endoscopic sinus surgery, he opted for conservative management with antihistamines, nasal rinses, and decongestants. Allergy and hematology evaluations were inconclusive, leading to a diagnosis of presumed IHES.

The patient was lost to follow-up and returned two years later with worsening sinusitis and new abdominal complaints. A CT scan of the abdomen and pelvis with contrast revealed retroperitoneal mesenteric lymphadenopathy with fat stranding, leading to a rheumatology referral; however, the antibody workup was negative. He was referred back to ENT, where he agreed to proceed with sinus surgery. Nasal biopsy findings from polyps showed acute and chronic lymphoplasmacytic inflammation with storiform fibrosis, eosinophils, and >10 IgG4 cells per high-power field, with an $IgG4^+/IgG^+$ ratio >40% suggestive of IgG4-related disease (IgG4-RD). No other common manifestations of IgG4-RD, such as salivary gland enlargement, autoimmune pancreatitis, orbital disease, or periaortitis/aortitis, were noted in this case. Treatment included a rapid prednisone taper and initiation of rituximab, with notable symptom improvement reported after four months.

DIFFERENTIAL DIAGNOSIS

In this case, the differential diagnoses were initially broad due to the complex presentation and progressive nature of the patient's symptoms. A wide range of conditions can mimic IgG4-RD and should be considered. Infectious causes such as tuberculosis, *Mycobacterium avium* complex, HIV, and syphilis can lead to retroperitoneal or mesenteric lymphadenopathy with fat stranding, often mimicking IgG4-RD.[1–4] However, our patient lacks systemic signs of infection and has no known risk factors, such as immunosuppression, animal exposure, or travel history. Also, imaging did not reveal necrotic or suppurative lymph nodes typical of some infections, and there was no histologic evidence of caseating granulomas, organisms, or infectious infiltrates on biopsy. Sarcoidosis is less likely given the absence of noncaseating granulomas and the lack of pulmonary or hilar lymph node involvement; although it remains a clinical mimicker of many conditions, current findings do not support the diagnosis.[5] Sjögren disease and lupus can have overlapping features with IgG4-RD but lack the characteristic biopsy findings.[6] Multicentric Castleman disease can present with lymphadenopathy and elevated IgG4 levels, but our patient lacks systemic inflammatory features or constitutional symptoms typically seen in it.[4]

Allergic fungal sinusitis is commonly associated with chronic rhinosinusitis and nasal polyposis, presenting with eosinophilia and elevated IgE levels.[7] However, negative fungal-specific serologic test results and the absence of typical ABPA markers made allergic fungal sinusitis unlikely.

Given the patient's elevated eosinophil count, chronic nasal congestion, and eosinophilic sinus disease, hypereosinophilic pneumonia was considered in the differential diagnosis, but the absence of pulmonary infiltrates and characteristic respiratory symptoms made it less likely.[8] Parasitic infections often present with an elevated eosinophil count, particularly during systemic migration phases. The patient was screened with specific parasitic serologic tests (*Toxocara*, *Strongyloides*) and a stool analysis, the results of which were negative, excluding parasitic etiology.[9,10]

Mastocytosis is characterized by elevated tryptase levels, which are key markers for systemic mast cell activation. Mastocytosis typically presents with skin lesions, gastrointestinal symptoms, and recurrent anaphylaxis-like episodes.[11] Although the tryptase level was elevated in this case, there were no associated systemic symptoms or skin manifestations. The bone marrow biopsy did not show signs of mast cell proliferation or aggregates, and no clonal markers of mastocytosis were identified.

Although hematologic malignancies, such as leukemia and lymphoma, are common causes of hypereosinophilia with systemic involvement, the workup including peripheral blood flow cytometry, bone marrow biopsy, and molecular testing for BCR-ABL, FIP1L1-PDGFRA, PDGFRB, and JAK2 V617F mutations was negative, effectively ruling out malignancy.[12] The elevated eosinophil count initially suggested HES, a group of disorders characterized by persistent eosinophilia and associated organ damage.[13] However, further hematologic evaluation confirmed a reactive, nonclonal process with no systemic involvement beyond the sinuses.

Eosinophilic granulomatosis with polyangiitis (EGPA), characterized by severe, persistent, and often worsening asthma, peripheral eosinophilia, and features of systemic vasculitis, including mononeuritis multiplex, pulmonary infiltrates, and potential cardiac or renal involvement, was initially suspected.[14] Although ANCA positivity is seen in approximately 30% to 35% of patients with EGPA, with myeloperoxidase–ANCA being the most common, it is not universally present.[15] In this case, the result of ANCA testing was negative, and there was no evidence of systemic vasculitis or organ involvement typically associated with EGPA, such as renal or cardiac involvement. Biopsy findings did not show the granulomas typically seen in EGPA, ruling out this diagnosis.

Although rare, drug-induced hypereosinophilia could be linked to ticagrelor, which may exacerbate conditions like IHES.[16] The patient's history of aspirin-exacerbated respiratory disease (AERD) could contribute to worsening respiratory symptoms, including sinus and asthma-related complaints. No direct link between ticagrelor and eosinophilia or sinusitis has been established, but medications should always be considered when symptoms evolve.

Rare differential diagnoses such as primary immunodeficiencies, including DOCK8 deficiency and hyper-IgE syndrome, should also be considered. DOCK8 deficiency, characterized by recurrent skin and viral infections, food allergies, and eczema-like changes, and hyper-IgE syndrome, marked by very high IgE levels, recurrent skin abscesses, pneumonia leading to bronchiectasis, and distinctive facial or skeletal features, were ruled out in this patient because he did not exhibit the typical signs of either.[17]

DISCUSSION

IgG4-RD is a systemic fibroinflammatory condition characterized by tissue infiltration with IgG4-positive plasma cells, storiform fibrosis, and often obliterative phlebitis. It can affect multiple organs, including the pancreas, bile ducts, salivary glands, and kidneys, presenting with mass-forming lesions and elevated serum IgG4 levels.[18] In the sinuses, it is recognized as a sclerosing condition, often associated with bone destruction and nasal septal damage.[19] The pathogenesis involves an autoimmune mechanism with $CD4^+$ T cells and B cells playing a central role.[20] AERD is a chronic inflammatory condition characterized by a clinical triad: asthma, chronic rhinosinusitis with nasal polyposis, and hypersensitivity to cyclooxygenase 1 inhibitors, such as aspirin and other nonsteroidal anti-inflammatory drugs. Diagnosis is often clinical, confirmed by an aspirin challenge, and treatment includes aspirin desensitization, leukotriene modifiers,

glucocorticoids, and biologics.[21–23] The coexistence of both conditions is a rare but recognized clinical phenomenon, previously described only once in the literature in a nonsmoking 61-year-old man with a history of nonsteroidal anti-inflammatory drug sensitivity who presented with a one-year history of left eye proptosis and chronic nasal symptoms and was ultimately diagnosed with concurrent AERD and IgG4-related sinusitis.[24]

Nasal polyp involvement in IgG4-RD is not typically a primary manifestation but is recognized in a subset of patients, often presenting as chronic rhinosinusitis with significant IgG4-positive plasma cell infiltration. Several studies have documented sinonasal involvement in IgG4-RD. For example, Ueno et al reported five cases of IgG4-RD with nasal mucosa and sinus involvement, in which patients initially presented with symptoms resembling allergic rhinitis or chronic sinusitis.[25] Also, Suzuki et al found that 56.5% of their patients with IgG4-RD exhibited significant IgG4-positive plasma cell infiltration in the nasal mucosa, emphasizing that nasal involvement is a notable feature in some patients with IgG4-RD.[26]

Type 2 inflammation is a shared immunopathologic feature between AERD and IgG4-RD. Both conditions are characterized by elevated levels of type 2 cytokines, such as interleukin-5 (IL-5) and IL-13, which drive eosinophilic inflammation and IgG4 production. In AERD, type 2 inflammation is marked by increased levels of IL-5 and IL-13, contributing to eosinophilic infiltration and severe nasal polyposis.[27,28] Additionally, IL-33 plays a significant role in amplifying type 2 immune responses in AERD, further promoting eosinophilic inflammation and mast cell activation.[29,30] Similarly, IgG4-RD involves a Th2-dominant immune response, with elevated IL-4, IL-10, and IL-13 levels, leading to tissue infiltration by IgG4-positive plasma cells and fibrosis.[31,32] The presence of IL-33 in IgG4-RD also suggests a shared pathway with AERD, as IL-33 can activate M2 macrophages and promote Th2 responses.[32,33] Buchheit et al found elevated levels of IgG4 in nasal polyp tissue from patients with AERD, whereas Engelhart et al demonstrated a significant association between AERD and elevated serum IgG4 levels, both supporting the link between these conditions.[34,35] Although there is evidence of overlap and coexistence between AERD and IgG4-RD, there is no current evidence to suggest that AERD is an early sign of IgG4-RD or that AERD evolves into IgG4-RD.

The comprehensive management strategy, spanning multiple specialties and tailored to evolving symptoms, underscores the complexity of treating IgG4-RD with multisystem involvement. Early intervention can halt disease progression, prevent organ damage, and alleviate chronic inflammation, making it essential to recognize IgG4-RD in patients with persistent eosinophilia and AERD features. This is crucial for appropriate management, as treatment strategies may need to address both the respiratory and systemic inflammatory components of the diseases.

FINAL DIAGNOSIS

Chronic rhinosinusitis with nasal polyps, complicated by asthma and aspirin intolerance (longstanding AERD or Samter's triad), coexisting with IgG4-related disease, as confirmed by subsequent nasal biopsy and laboratory results.

AUTHOR CONTRIBUTIONS

Both 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 Sondhi confirms that the 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. Greco P, Vaglio A, Corradi D, et al. Tuberculosis as a trigger of retroperitoneal fibrosis. Clin Infect Dis 2005;41(7):e72–e75.
2. Liu TT, Weng SW, Wang MC, et al. Nontuberculous mycobacterial infection with concurrent IgG4-related lymphadenopathy. APMIS 2016;124(3):216–220.
3. Yu HT, Lee CH, Huang SC, et al. Unsuspected human immunodeficiency virus infection presenting as immunoglobulin G4-related lymphadenopathy: a case report. Int J STD AIDS 2018;29(1):92–95.
4. Cheuk W, Bledsoe JR. IgG4-related lymphadenopathy. Semin Diagn Pathol 2024;41(2):108–115.
5. Huynh KN, Kong MJ, Nguyen BD. Anatomic and functional imaging of immunoglobulin G4-related disease and its mimics. Radiographics 2023;43(3):e220097.
6. Fragoulis GE, Zampeli E, Moutsopoulos HM. IgG4-related sialadenitis and Sjögren's syndrome. Oral Dis 2017;23(2):152–156.
7. Dykewicz MS, Rodrigues JM, Slavin RG. Allergic fungal rhinosinusitis. J Allergy Clin Immunol 2018;142(2):341–351.
8. Cottin V. Eosinophilic lung diseases. Immunol Allergy Clin North Am 2023;43(2):289–322.
9. Yoon SY, Baek S, Park SY, et al. Clinical course and treatment outcomes of toxocariasis-related eosinophilic disorder. Medicine (Baltimore) 2018;97(37):e12361.
10. Czeresnia JM, Weiss LM. Strongyloides stercoralis. Lung 2022; 200(2):141–148.
11. Gülen T, Hägglund H, Dahlén B, et al. Mastocytosis: the puzzling clinical spectrum and challenging diagnostic aspects of an enigmatic disease. J Intern Med 2016;279(3):211–228.
12. Gerds AT, Gotlib J, Bose P, et al. Myeloid/lymphoid neoplasms with eosinophilia and TK fusion genes, Version 3.2021, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 2020; 18(9):1248–1269.
13. Shomali W, Gotlib J. World Health Organization and International Consensus Classification of eosinophilic disorders: 2024 update on diagnosis, risk stratification, and management. Am J Hematol 2024; 99(5):946–968.
14. Grayson PC, Ponte C, Suppiah R, et al; DCVAS Study Group. 2022 American College of Rheumatology/European Alliance of Associations for Rheumatology classification criteria for eosinophilic granulomatosis with polyangiitis. Arthritis Rheumatol 2022;74(3):386–392.

15. Moiseev S, Bossuyt X, Arimura Y, et al; European EGPA Study Group. International consensus on antineutrophil cytoplasm antibodies testing in eosinophilic granulomatosis with polyangiitis. Am J Respir Crit Care Med 2020;202(10):1360–1372.
16. Das C, Dharmkare D, Mani S, et al. Drug rash with eosinophilia and systemic symptoms (DRESS) syndrome following ticagrelor administration leading to serious complications. Med Res Arch 2024;12(2):2.
17. Biggs CM, Keles S, Chatila TA. DOCK8 deficiency: insights into pathophysiology, clinical features and management. Clin Immunol 2017; 181:75–82.
18. Maritati F, Peyronel F, Vaglio A. IgG4-related disease: a clinical perspective. Rheumatology (Oxford) 2020;59(suppl 3):iii123–iii131.
19. Kaur K, Kakkar A, Manchanda S, et al. Sinonasal IgG4-related disease: a rare and emerging entity broadening the differential diagnosis in the sinonasal universe. Eur Arch Oto-Rhino-Laryngol 2021;278(8): 2883–2890.
20. Perugino CA, Stone JH. IgG4-related disease: an update on pathophysiology and implications for clinical care. Nat Rev Rheumatol 2020;16(12):702–714.
21. Stevens WW, Jerschow E, Baptist AP, et al. The role of aspirin desensitization followed by oral aspirin therapy in managing patients with aspirin-exacerbated respiratory disease: a work group report from the Rhinitis, Rhinosinusitis and Ocular Allergy Committee of the American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol 2021;147(3):827–844.
22. Luskin K, Thakrar H, White A. Nasal polyposis and aspirin-exacerbated respiratory disease. Immunol Allergy Clin North Am 2020;40(2):329–343.
23. Ledford DK, Lockey RF. Aspirin or nonsteroidal anti-inflammatory drug-exacerbated chronic rhinosinusitis. J Allergy Clin Immunol Pract 2016;4(4):590–598.
24. Johal K, Welch K, Peters A. Immunoglobulin G4 sinusitis in association with aspirin-exacerbated respiratory disease. Am J Rhinol Allergy 2017;31(5):302–304.
25. Ueno M, Nakano K, Miyagawa I, et al. Five cases of IgG4-related disease with nasal mucosa and sinus involvement. Intern Med 2020; 59(15):1905–1911.
26. Suzuki M, Nakamaru Y, Akazawa S, et al. Nasal manifestations of immunoglobulin G4-related disease. Laryngoscope 2013;123(4): 829–834.
27. Scott WC, Cahill KN, Milne GL, et al. Inflammatory heterogeneity in aspirin-exacerbated respiratory disease. J Allergy Clin Immunol 2021;147(4):1318–1328.e5.
28. Bangert C, Villazala-Merino S, Fahrenberger M, et al. Comprehensive analysis of nasal polyps reveals a more pronounced type 2 transcriptomic profile of epithelial cells and mast cells in aspirin-exacerbated respiratory disease. Front Immunol 2022;13:850494.
29. Liu T, Barrett NA, Kanaoka Y, et al. Type 2 cysteinyl leukotriene receptors drive IL-33-dependent type 2 immunopathology and aspirin sensitivity. J Immunol 2018;200(3):915–927.
30. Wangberg H, White AA. Aspirin-exacerbated respiratory disease. Curr Opin Immunol 2020;66:9–13.
31. Kasashima S, Kawashima A, Zen Y, et al. Upregulated interleukins (IL-6, IL-10, and IL-13) in immunoglobulin G4-related aortic aneurysm patients. J Vasc Surg 2018;67(4):1248–1262.
32. Michailidou D, Schwartz DM, Mustelin T, et al. Allergic aspects of IgG4-related disease: implications for pathogenesis and therapy. Front Immunol 2021;12:693192.
33. Ishiguro N, Moriyama M, Furusho K, et al. Activated M2 macrophages contribute to the pathogenesis of IgG4-related disease via Toll-like receptor 7/interleukin-33 signaling. Arthritis Rheumatol 2020;72(1): 166–178.
34. Buchheit KM, Dwyer DF, Ordovas-Montanes J, et al. IL-5Rα marks nasal polyp IgG4- and IgE-expressing cells in aspirin-exacerbated respiratory disease. J Allergy Clin Immunol 2020;145(6):1574–1584.
35. Engelhart S, Glynn RJ, Schur PH. Disease associations with isolated elevations of each of the four IgG subclasses. Semin Arthritis Rheum 2017;47(2):276–280.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 188–196
DOI 10.1002/acr.25609

Trends in Dermatopolymyositis Mortality, 1999–2022: A Nationwide Population-Based Study, United States

Elizabeth Matz and Ram R. Singh

Objective. We evaluated trends in dermatopolymyositis (DPM) mortality relative to all-cause mortality in the United States from 1999 to 2022.

Methods. We used the Centers for Disease Control and Prevention's databases (Multiple Causes of Death for 1999–2020 and Provisional Mortality Statistics for 2021 and 2022) to obtain death counts for DPM and non-DPM (all causes other than DPM). We calculated age-standardized mortality rates (ASMRs) for both groups and computed the ratio of DPM-ASMR to non–DPM-ASMR for each of the 24 years. We performed joinpoint regression analysis to estimate annual percent change (APC) in DPM-ASMRs and non–DPM-ASMRs and in the DPM-ASMR to non–DPM-ASMR ratios overall and by sex, age, and race and ethnicity.

Results. There were 12,882 DPM and 63,549,485 non-DPM deaths during 1999–2022. Mortality decreased at a higher APC (–3.8% [95% confidence interval (CI) –4.3% to –3.4%]) for DPM than for non-DPM (–1.2% [95% CI –1.5% to –0.9%]) until the start of the COVID-19 pandemic, when it increased for both at similar APCs. Consequently, the ratios of DPM-ASMRs to non–DPM-ASMRs decreased over these 24 years in all subgroups. The DPM-ASMR to non–DPM-ASMR ratios were higher in females than males and in younger individuals (≤64 years) than those aged ≥65 years. The odds of premature death were higher for DPM than for non-DPM. Individuals of non-Hispanic Black, Hispanic, and non-Hispanic other race and ethnicity had higher DPM-ASMR to non–DPM-ASMR ratios than White individuals.

Conclusion. DPM mortality decreased at a higher rate than all-cause mortality until the pandemic, when it proportionately increased for both DPM and all causes. Females, younger individuals, and individuals of Black, Hispanic, and non-Hispanic other race and ethnicity had higher DPM mortality relative to all-cause mortality. Findings highlight the need for improved screening, earlier intervention, and targeted efforts to address racial and ethnic disparities in DPM outcomes.

INTRODUCTION

Dermatopolymyositis (DPM) comprises chronic autoimmune inflammatory diseases that involve the muscles, skin, lungs, esophagus, heart, joints, and other organs.[1–4] DPM can lead to premature death due to active disease manifestations, including muscle weakness, lung and heart disease, adverse effects from treatment (including infections), and cancers.[5,6] The five-year mortality rates for DPM were as high as 23% to 73% before the introduction of glucocorticoids for its treatment.[6] Survival for DPM has improved since then, but initial treatment with high-dose glucocorticoids was also associated with lower survival[7] from 1997 to 2003. DPM outcomes have improved further with the fine-tuning of glucocorticoids and other immunosuppressive regimens, early recognition of high-risk disease features, and increasing use of intravenous Ig.[8,9] Assessing the impact of these advances on DPM mortality burden at the population level will further help improve DPM outcomes.

In a Michigan cohort of 160 patients, male sex and race and ethnicity other than White race were associated with lower survival.[7] However, mortality was not influenced by sex or race and ethnicity in a retrospective analysis of 831 patients with DPM from 2006 through 2014 at the Johns Hopkins Myositis Center.[10] Although the cohort-based studies provide useful information,

Ms Matz is a recipient of the 2025 Medical and Graduate Student Preceptorship from the Rheumatology Research Foundation.

Elizabeth Matz, BS, Ram R. Singh, MBBS, MD: University of California at Los Angeles.

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.25609).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25609.

Address correspondence via email to Ram R. Singh, MBBS, MD, at rrsingh@mednet.ucla.edu.

Submitted for publication January 8, 2025; accepted in revised form June 26, 2025.

SIGNIFICANCE & INNOVATIONS

- Using the largest dermatopolymyositis (DPM) death counts (12,882) to date, this study reports decreasing DPM mortality across the United States over the last quarter-century.
- This is the first study to assess DPM mortality in the context of all-cause mortality, reporting that DPM mortality decreased at a higher rate than all-cause mortality over a quarter-century.
- This is the first population-based study to compare DPM mortality before and during the COVID-19 pandemic, and it shows that the increase in DPM mortality during the pandemic was proportionate to the increase in all-cause mortality.
- Females, younger individuals, and individuals of non-Hispanic Black, Hispanic, and non-Hispanic other race and ethnicity had higher DPM mortality relative to all-cause mortality, and the odds of premature death were higher for DPM than for all other causes of death.

they may not represent the burden of disease at the population level.

Population-based data on DPM mortality are scarce in the United States. An early study of 1,986 DPM deaths from 1968 through 1978 revealed a higher age-adjusted mortality rate in race and ethnicity other than White females than in White females and both race and ethnicity other than White and White males.[11] Another study[12] retrieved DPM death counts from the US national mortality database from 1981 to 2020, but the association of DPM mortality with age and race and ethnicity was not explored. A recent population-based study of dermatomyositis in Olmsted County reported nine deaths from 1995 to 2019, giving the standardized mortality ratio of 2.0 (95% confidence interval [CI] 0.9–3.7).[13] However, a small sample size, particularly of individuals of Native American, Asian, Black, Hispanic, and Pacific Islander race and ethnicity or of male sex, may have limited its utility in assessing DPM burden in the US population.[11,13]

Here, we analyzed all recorded deaths in the United States from 1999 through 2022 to evaluate temporal trends in DPM mortality overall and by sex, age, and race and ethnicity. To evaluate DPM mortality in the context of changes in general population mortality, we calculated the ratios of age-standardized mortality rates (ASMRs) for DPM to ASMRs for all causes other than DPM (non-DPM).

METHODS

Data source. Annual mortality data were obtained using the Centers for Disease Control and Prevention's (CDC's) Wide-ranging Online Data for Epidemiologic Research (WONDER) databases.[14] Death counts for DPM were retrieved from the Multiple Causes of Death (MCD) 1999–2020 database using *International Classification of Diseases, Tenth Revision* codes (Supplemental Table S1). Data for 2021 and 2022 were retrieved from MCD Provisional Mortality Statistics 2018–2022 and merged with the 1999–2020 data. Average population data were obtained from the US Census Bureau.

ASMR. Age-specific crude mortality rates for DPM and non-DPM were calculated as the number of deaths in each year divided by the number of persons in the US general population in that same year. The US population from the year 2000 was used to calculate ASMR per million, with 95% CIs,[15] overall as well as separately for males and females, for each of the four racial and ethnic groups, and for three age groups (≤44, 45–64, and ≥65 years) for each year from 1999 through 2022. Racial and ethnic groups included Hispanic, non-Hispanic Black, non-Hispanic White, and non-Hispanic other (Asian, Pacific Islander, American Indian, or Alaska Native). We then computed the ratios of DPM-ASMR to non–DPM-ASMR for each year from 1999 through 2022 overall and separately for each sex, age, and racial and ethnic subgroup. We also conducted a post hoc analysis comparing odds of premature death in DPM with two other autoimmune diseases: systemic lupus erythematosus (SLE) and systemic sclerosis (SSc).

Statistical analysis. The joinpoint regression model was used to evaluate mortality trends overall and by sex, age, and race and ethnicity. Joinpoint regression identifies changes in trend data by fitting a set of joinpoints—the calendar years at which the change in the slope (of ASMR or ratio) is statistically significant—over the entire period. The model then computes the slope (year-to-year percentage change in annual ASMR or ratio) and the 95% CI over each linear trend segment between adjacent joinpoints.[16,17]

The joinpoint regression model assumes that its regression mean function is piecewise linear and that the linear segments are continuously connected at the joinpoints. The program allows one of three options for handling heteroscedastic errors: constant variance (homoscedasticity), SE, or Poisson variance. The program uses one of two methods for model fitting: the grid search method or the Hudson method. The grid search method identifies a discrete number of locations for testing for changes in slope, and the Hudson method allows for continuous model fitting. For this study, we provided the SE and used the grid search method.

The Joinpoint Regression Program (version 4.2.0.2; National Cancer Institute) uses a permutation test to find the optimal number of joinpoints. For greater consistency in the permutation test *P* values, the Joinpoint Regression Program runs at least 4,499 permutations to select the optimal number of linear segments to fit the model. Because fitting all 4,499! (factorial) possible permutations would be computationally intensive, the program uses a

Monte Carlo simulation to conduct significance tests on a sample of the 4,499! permutations, which are adjusted using the Bonferroni correction to reduce the likelihood of false positive results.[18] The study is exempt from ethics committee approval requirement.

RESULTS

From 1999 through 2022, DPM was recorded as the underlying or a contributing cause of death in 12,882 decedents. Females (64.7%), non-Hispanic Black individuals (17.4%), and those in the 45 to 64 years age group (28.0%) were overrepresented among DPM decedents relative to their representation among non-DPM (all causes other than DPM) decedents (Table 1).

Trends in mortality: DPM versus non-DPM (all cause), 1999–2022. Mortality rates decreased for both DPM and non-DPM from 1999 through 2022 but at a higher pace for DPM (annual percent change [APC] –3.8 [95% CI –4.3 to –3.4]) than for non-DPM (APC –1.2 [95% CI –1.5 to –0.9]) until the start of the COVID-19 pandemic (Figure 1). The cumulative percent change over the entire 24-year period was –46.8% for DPM but only –4.8% for non-DPM (Supplemental Table S2). During the pandemic, mortality rates significantly increased for both DPM and non-DPM at a similar APC. Consequently, the ratio of DPM to non-DPM mortality rates continuously decreased throughout the 24-year period. This suggests that there was no disproportionate increase in DPM-specific mortality during the COVID-19 pandemic.

Trends in mortality by sex, age, and race and ethnicity: DPM versus non-DPM, 1999–2022. Females had higher DPM mortality rates than males; the converse was true for non-DPM mortality rates, which were higher for males (Figure 2). There were no significant differences in the cumulative percent changes for both DPM and non-DPM deaths by sex (Supplemental Table S2).

Similar mortality trends were observed at different age groups for DPM and non-DPM, with the highest ASMRs in the ≥65 years age group for both DPM and non-DPM (Figure 3). However, the ratio of DPM-ASMRs to non–DPM-ASMRs was higher in the 45 to 64 and ≤44 years age groups than in those aged ≥65 years, suggesting a disproportionately higher mortality rate at younger age groups for DPM than for non-DPM. The odds of premature death (<65 years) from DPM were significantly increased compared with non-DPM (odds ratio [OR] 1.6; 95% CI 1.5–1.6), which were less than in SSc (OR 1.9; 95% CI 1.9–2.0) and SLE (OR 4.3; 95% CI 4.2–4.4) (Table 2).

The cumulative decreases in DPM mortality rates were also less in younger age groups than in older age groups (Supplemental Table S2). Notably, there was a cumulative increase of 21.1% in non–DPM-ASMRs for the ≤44 years age group, suggesting a higher premature all-cause mortality rate in 2022 than in 1999.

Mortality rates in non-Hispanic Black persons were higher than those in other racial and ethnic groups for both DPM and non-DPM causes (Figure 4). The ratio of DPM-ASMRs to non–DPM-ASMRs was higher in individuals of non-Hispanic Black, Hispanic, and non-Hispanic other race and ethnicity than in non-Hispanic White individuals. The 24-year cumulative decrease in DPM mortality rates was more than 50% for non-Hispanic Black and non-Hispanic White persons but only 8.7% for non-Hispanic Asian, Pacific Islander, American Indian, and Alaska Native individuals and 34.5% for Hispanic individuals (Supplemental Table S2).

Table 1. Demographic characteristics, 1999–2022*

Characteristics	DPM deaths[a]		Non-DPM deaths[a]		Average population[b]	
	No.	%	No.	%	No.	%
Total	12,882		63,549,485		7,410,144,137	
Sex						
Female	8,339	64.7	31,504,793	49.6	3,764,021,810	50.8
Male	4,543	35.3	32,042,977	50.4	3,646,122,327	49.2
Age						
≤44 y	1,024	8.0	4,837,629	7.6	4,508,987,499	60.8
45–64 y	3,610	28.0	12,049,566	19.0	1,860,984,067	25.1
≥65 y	8,248	64.0	46,657,324	73.4	1,040,172,571	14.0
Not stated	0	0	4,966	0.0	–	–
Race and ethnicity						
NH White	8,903	69.1	49,961,434	78.6	4,787,848,120	64.6
NH Black	2,240	17.4	7,585,215	11.9	947,648,882	12.8
Hispanic	1,128	8.8	4,028,734	6.3	1,139,927,382	15.4
NH A/PI/AN/AI	591	4.6	1,764,980	2.8	456,491,367	6.2
Not stated	20	0.2	209,122	0.3	78,228,386	1.1

* A, Asian; AI, American Indian; AN, Alaska Native; Black, Black or African American; DPM, dermatopolymyositis; Hispanic, Hispanic or Latino; NH, non-Hispanic; non-DPM, all-cause deaths other than DPM; PI, Pacific Islander.
[a] Number (percentage) of deaths from all 50 states and the District of Columbia as of September 14, 2023, in all categories except for total and sex in non-DPM (63,547,770 as of September 7, 2023).
[b] Mean annual population (percentage) derived from US Census Bureau files.

DISCUSSION

Analysis of all recorded deaths attributed to DPM in the United States from 1999 through 2022 revealed that mortality rates have decreased at a higher pace for DPM than for all-cause (non-DPM) deaths. There were significant differences between DPM and all-cause mortality by sex, age, and race and ethnicity. DPM was associated with higher odds of premature death compared to all-cause deaths. During the pandemic, mortality rates increased for both DPM and all causes at similar APCs; thus, the ratios of DPM to all-cause mortality rates continuously decreased throughout the 24-year period in all sex, age, and racial and ethnic groups, suggesting no disproportionate increase in DPM-associated mortality during the pandemic.

A few population-based studies have examined the changes in DPM mortality over time. An early study[11] reported increased DPM mortality rates in the United States from 1968 to 1978. A 1999–2014 study[19] of 303 DPM deaths from British Columbia, Canada, revealed a 2.4- to 5.5-fold increased risk of DPM death from 1997 to 2005 compared to a 2.0- to 4.1-fold increased risk of DPM death from 2006 to 2014. A 1999–2014 United Kingdom study of 114 DPM deaths did not report a significant improvement in excess mortality risk in DPM in recent years (2007–2014 vs 1999–2006).[20] However, we report that the mortality rates decreased at higher annual percent rates for DPM than for all-cause deaths from 1999 through 2018 in the US population.

The improving DPM mortality rate over the last two decades is likely attributable to a combination of factors (Figure 1, bottom panel),[2,3,8,21] including: (1) the 2003 European Neuromuscular Centre criteria, which updated the Bohan and Peter criteria by adding more muscle pathology details as well as incorporating amyopathic dermatomyositis[22]; (2) the increasing awareness

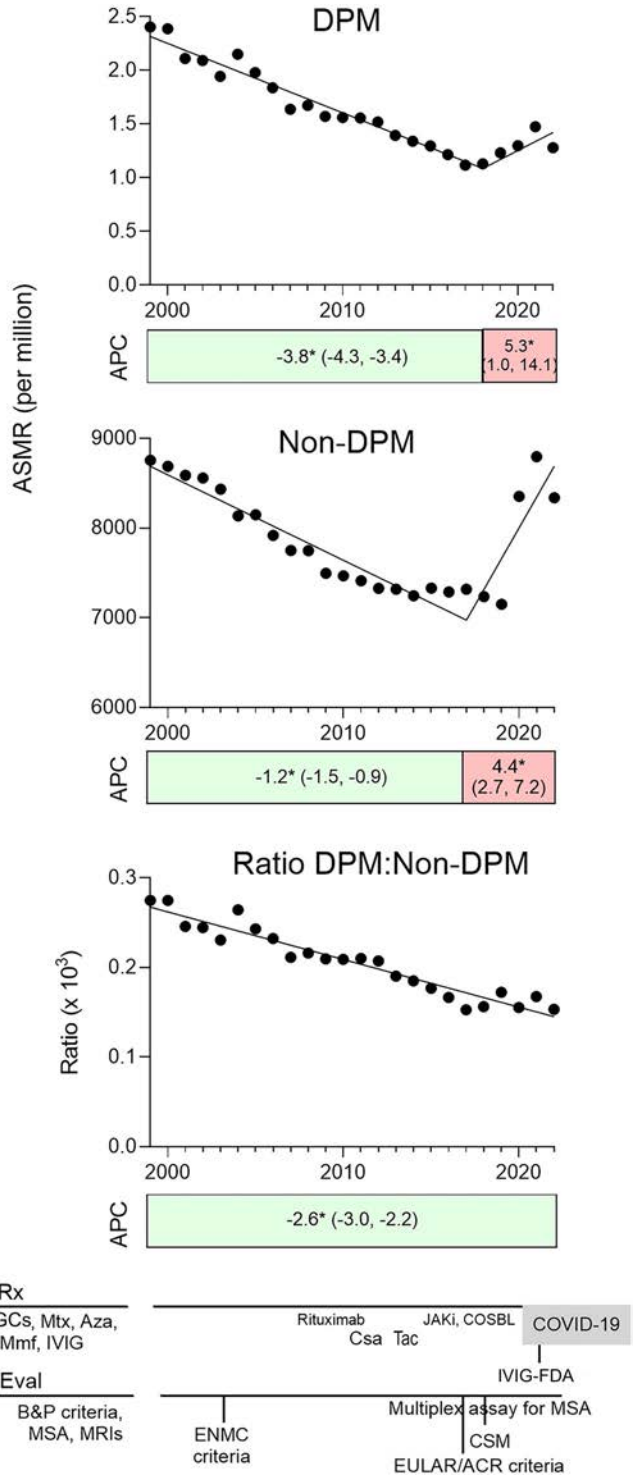


Figure 1. Legend on next page.

Figure 1. Trends in mortality rates for DPM and non-DPM causes and ratio of DPM to non-DPM mortality rates, 1999–2022. Results are expressed as the ASMR per million persons for DPM (top panel) and non-DPM (middle panel) and the ratio of DPM to non-DPM mortality rates (lower panel). Data are displayed per calendar year of death, with lines fitted on the basis of joinpoint trend analysis. The annual number of deaths ranged from 428 to 664 for DPM and from 2,391,374 to 3,464,216 for non-DPM. The APC for each trend is presented as stack bars below each panel. Each stack is segmented at the year in which the change in slope is statistically significant and is aligned with the trend line. Numbers in each stack denote the APC (95% confidence interval). The red-shaded stacks indicate an increasing trend, and the green-shaded stacks represent a decreasing trend. *$P < 0.05$ for slope change. (Bottom panel) DPM-ASMR is shown in relation to major therapeutic developments (Rx), advances in evaluation (Eval) of DPM, and the COVID-19 pandemic. ACR, American College of Rheumatology; APC, annual percent change; ASMR age-standardized mortality rate; Aza, azathioprine; B&P criteria, Bohan and Peter classification criteria; COSBL, T cell costimulation blockade; Csa, cyclosporine A; CSM, core set measures; DPM, dermatopolymyositis; ENMC, European Neuromuscular Centre; GCs, glucocorticoids; IVIG, intravenous Ig; IVIG-FDA, IVIG treatment approved by the Food and Drug Administration; JAKi, JAK inhibitors; Mmf, mycophenolate mofetil; MRI, magnetic resonance imaging; MSA, myositis-specific autoantibodies; Mtx, methotrexate; non-DPM, all-cause deaths other than DPM; Rx, therapeutic advances; Tac, tacrolimus.

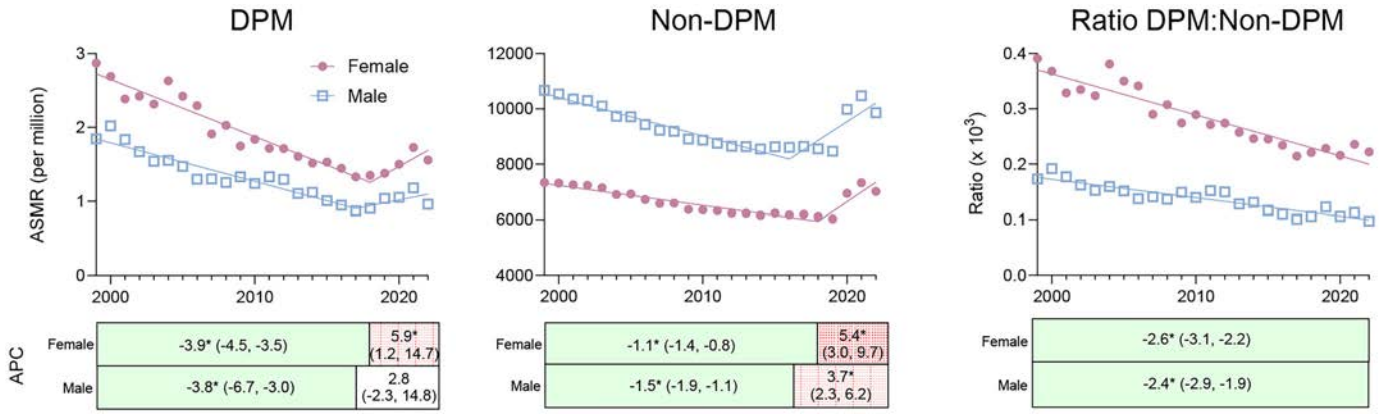


Figure 2. Trends in mortality rates for DPM and non-DPM and in the ratio of DPM mortality rates to non-DPM mortality rates by sex, 1999–2022. Results are expressed as the ASMR per million persons for DPM (left panel) and non-DPM (middle panel) and the ratio of DPM to non-DPM mortality rates (right panel). Data are displayed per calendar year of death, with lines fitted on the basis of joinpoint trend analysis. The annual number of DPM deaths ranged from 277 to 445 for females and from 151 to 240 for males. The APC for each trend for each subpopulation is presented as stack bars below each panel. Each stack is segmented at the year in which the change in slope is statistically significant and is aligned with the trend line. Numbers in each stack denote the APC (95% confidence interval). The red-shaded stacks indicate an increasing trend, unshaded stacks represent a nonsignificant trend, and the green-shaded stacks represent a decreasing trend. *$P < 0.05$ for slope change. APC, annual percent change; ASMR, age-standardized mortality rate; DPM, dermatopolymyositis; non-DPM, all-cause deaths other than DPM. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25609/abstract.

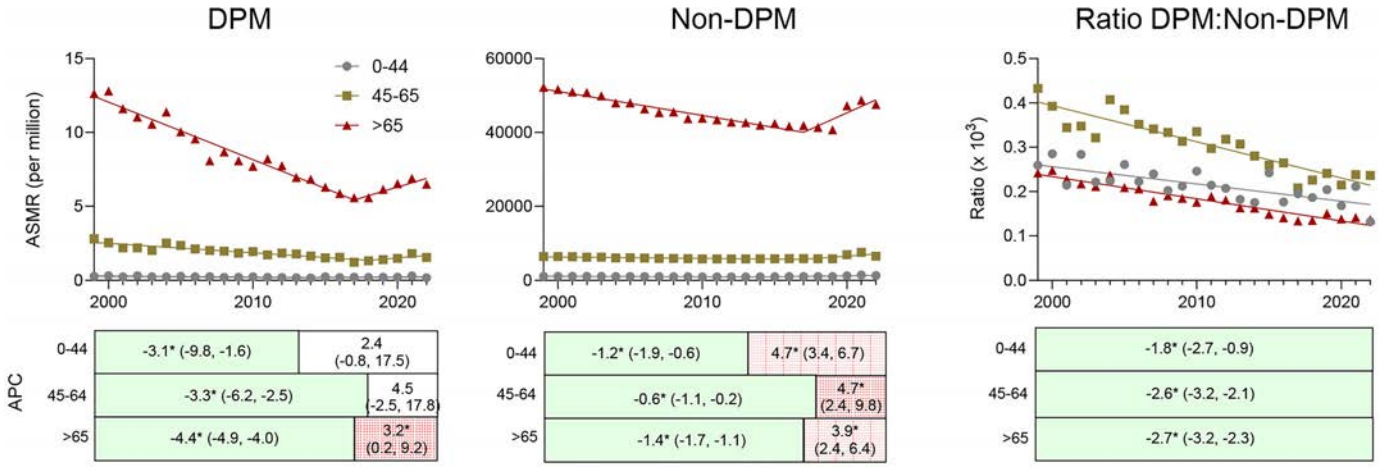


Figure 3. Trends in mortality rates for DPM and non-DPM and in the ratio of DPM mortality rates to non-DPM mortality rates by age, 1999–2022. Results are expressed as the ASMR per million persons for DPM (left panel) and non-DPM (middle panel) and the ratio of DPM to non-DPM mortality rates (right panel). Data are displayed per calendar year of death, with lines fitted on the basis of joinpoint trend analysis. The annual number of DPM deaths ranged from 33 to 57 for those aged ≤44 years, from 112 to 183 among those aged 45 to 64 years, and from 275 to 447 among those aged ≥65 years. The APC for each trend for each subpopulation is presented as stack bars below each panel. Each stack is segmented at the year in which the change in slope is statistically significant and is aligned with the trend line. Numbers in each stack denote the APC (95% confidence interval). The red-shaded stacks indicate an increasing trend, unshaded stacks represent a nonsignificant trend, and the green-shaded stacks represent a decreasing trend. *$P < 0.05$ for slope change. APC, annual percent change; ASMR, age-standardized mortality rate; DPM, dermatopolymyositis; non-DPM, all-cause deaths other than DPM. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25609/abstract.

and earlier recognition of interstitial lung disease as a manifestation of DPM[23,24]; (3) the recognition of myositis-specific antibodies as valuable tools in diagnosing and understanding DPM[25]; (4) the increased availability and reliability of myositis antibody testing, including the development of multiplex assays for testing[25]; (5) the development of DPM core set measures[26] and the 2017 EULAR/American College of Rheumatology criteria[42]; (6) earlier use of disease-modifying antirheumatic drugs, for example, methotrexate, azathioprine, and mycophenolate mofetil, with a more rapid tapering of glucocorticoids[21]; (7) increasing use of intravenous Ig[9,27]; (8) introduction of rituximab and the calcineurin inhibitors tacrolimus and cyclosporine A[21,28–31]; and (9) use of JAK inhibitors and costimulation blockers in patients refractory to current treatments.[21,32] Better management of comorbidities

Table 2. Odds of premature death, 1999–2022*

Age, y	DPM, n (%)	Non-DPM, n (%)	OR[a] (95% CI)	SSc, n (%)	Non-SSc, n (%)	OR[a] (95% CI)	SLE, n (%)	Non-SLE, n (%)	OR[a] (95% CI)
<65	4,634 (36.0)	16,887,195 (26.6)	1.6[b] (1.5–1.6)	17,654 (41.0)	16,885,593 (26.6)	1.9[b] (1.9–2.0)	32,009 (60.8)	16,884,595 (26.6)	4.3[b] (4.2–4.4)
≥65	8,248 (64.0)	46,657,324 (73.4)	1 (ref)	25,392 (59.0)	46,655,625 (73.4)	1 (ref)	20,659 (39.2)	46,656,511 (73.4)	1 (ref)

* CI, confidence interval; DPM, dermatopolymyositis; non-DPM, all causes other than DPM; OR, odds ratio; ref, reference; SLE, systemic lupus erythematosus; SSc, systemic sclerosis.
[a] Odds of premature death (≤65 y) from DPM, SSc, or SLE relative to all-cause (non-DPM, non-SSc, or non-SLE, respectively) deaths.
[b] $P < 0.0001$.

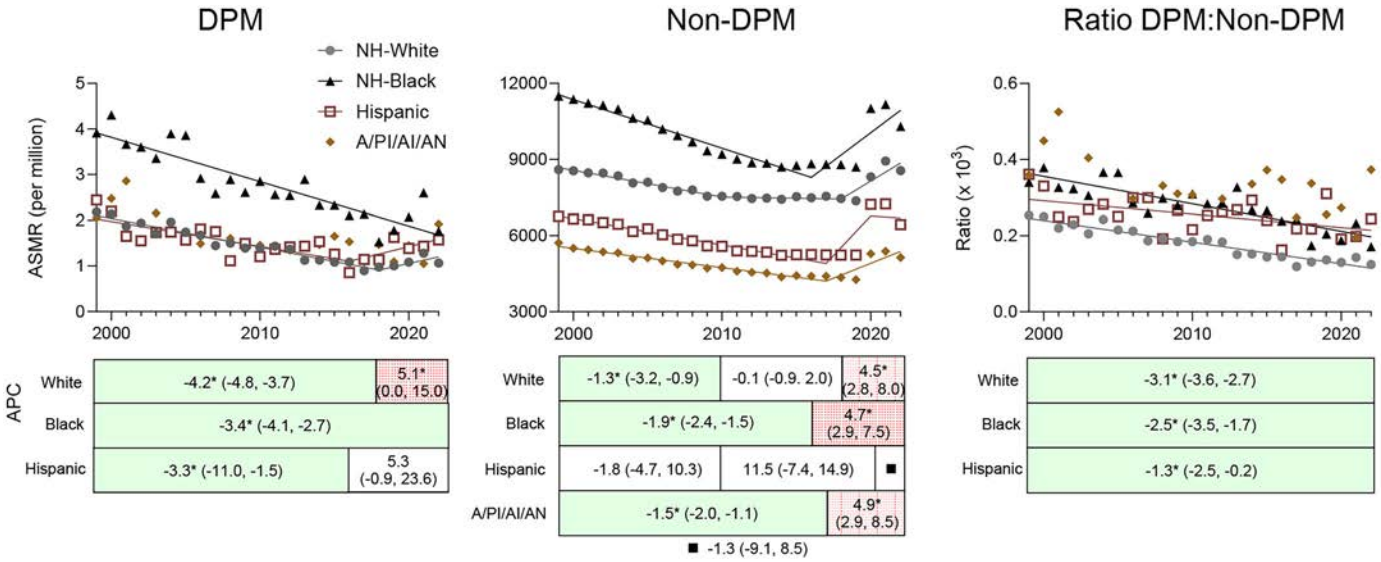


Figure 4. Trends in mortality rates for DPM and non-DPM and in the ratio of DPM mortality rates to non-DPM mortality rates by race and ethnicity, 1999–2022. Results are expressed as the ASMR per million persons for DPM (left panel) and non-DPM (middle panel) and the ratio of DPM to non-DPM mortality rates (right panel). Data are displayed per calendar year of death, with lines fitted on the basis of joinpoint trend analysis. The annual number of DPM deaths ranged from 270 to 491 for NH White individuals, from 61 to 116 for NH Black individuals, from 33 to 76 for Hispanic individuals, and from 14 to 46 for NH A/PI/AI/AN individuals. The APC for each trend for each subpopulation is presented as stack bars below each panel. Each stack is segmented at the year in which the change in slope is statistically significant and is aligned with the trend line. Numbers in each stack denote the APC (95% confidence interval). The red-shaded stacks indicate an increasing trend, unshaded stacks represent a nonsignificant trend, and the green-shaded stacks represent a decreasing trend. *$P < 0.05$ for slope change. Stack bars are not shown for NH A/PI/AI/AN under DPM and ratio trendline curves because the annual death counts were low in this group. A, Asian; AI, American Indian; AN, Alaska Native; APC, annual percent change; ASMR, age-standardized mortality rate; Black, Black or African American; DPM, dermatopolymyositis; Hispanic, Hispanic or Latino; NH, non-Hispanic; non-DPM, all-cause deaths other than DPM; PI, Pacific Islander. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25609/abstract.

including ischemic heart disease and malignancy, which are common causes of death in DPM,[5,6] could also have contributed to the decreasing DPM mortality rate. The decreasing mortality rate is not likely due to decreasing prevalence, as a recent estimate from Olmsted County suggested a relatively higher prevalence of dermatomyositis (13 per 100,000) than previous prevalence estimates,[13] which ranged from 1.2 to 9.2 per 100,000.

Mortality was not influenced by sex in the Johns Hopkins Myositis cohort,[10] whereas an association between male sex and lower survival was reported in the Michigan cohort.[7] Consistent with the known higher prevalence of DPM in females,[33] DPM mortality rates were consistently 1.6-fold higher in females than males: 2.9 versus 1.8 (1999) and 1.6 versus 1.0 (2022), respectively (Supplemental Table S2). Because DPM is two to three times more prevalent in females than males, with an adjusted OR of 2.52 (95% CI 1.08–3.99),[33] and mortality rates were only 1.6-fold higher in females, it is likely that males would have a higher case fatality rate than females when adjusted for prevalence.

We also found that DPM mortality rates were higher in non-Hispanic Black individuals than in non-Hispanic White individuals throughout the 24-year period. In Hispanic individuals and those of non-Hispanic other race and ethnicity (Asian, Pacific Islander, American Indian, Alaska Native), DPM mortality rates were not

different relative to non-Hispanic White persons in 1999. However, Black and White individuals had a more than 50% cumulative decrease in DPM mortality rates over the 24-year period, whereas those of non-Hispanic other race and ethnicity had only an 8.7% cumulative decrease. Thus, DPM mortality rates were significantly higher in all individuals of non-Hispanic Black, Hispanic, and non-Hispanic other race and ethnicity relative to non-Hispanic White individuals in 2022. Previous studies have reported either no influence of race and ethnicity on DPM mortality[10] or an association of race and ethnicity other than White race with lower survival from DPM.[7] Poorer clinical outcomes have also been reported in Black and minority non-Black children with dermatomyositis compared with White children.[34] Multiple factors likely contribute to these disparities. For example, diagnostic delay of DPM, which has been linked to the lack of access to specialists, has extensive impacts on quality of life with this disease.[35] Significantly prolonged time to diagnosis and treatment of dermatomyositis occurs in all other races when compared with White patients, suggesting a lack of dermatology education of race and ethnicity other than White skin in DPM detection.[36] Another study reported that negative beliefs about biologic therapies were significantly more prevalent (60%) in patients of other ethnicities compared to the White cohort (35%),[37] which could have contributed to higher mortality in individuals of other races and ethnicities compared to non-Hispanic White individuals with DPM. Further studies are needed to understand reasons underlying racial and ethnic disparities in DPM mortality.

We found the highest mortality rates in those aged ≥65 years for both DPM and all-cause deaths, which is consistent with a previous report that increased age at diagnosis was associated with lower survival.[7] However, we found that the ratio of DPM to all-cause mortality rates was higher in the 45 to 64 years and ≤44 years age groups than in those aged ≥65 years, suggesting higher premature mortality for DPM than for all-cause deaths. Consistently, the OR for premature death (≤64 years) was significantly higher for DPM relative to all-cause deaths. Studies are underway to identify causes of premature death in DPM. Our preliminary analysis suggests that myositis and interstitial lung disease were more commonly recorded as the underlying causes of death in DPM at younger ages than ≥65 years.

Notably, the all-cause (non-DPM) mortality rate was not substantially lower in 2022 than in 1999. In fact, the all-cause mortality rate at younger ages (≤44 years) has steeply increased by 4.7% each year since 2014, with a cumulative increase of 21.1% over 24 years. This worrisome increase in premature mortality in the general population has been attributed to increases in firearm-related deaths and drug overdose and poisoning among children and adolescents.[38]

About half of US rheumatologists in 2020 believed that patients with autoimmune and rheumatic diseases were at a higher risk for COVID-19, and as a result, there was a decreased use of biologics to treat these conditions during the pandemic,[39] which could have increased DPM mortality. However, we found that the overall DPM mortality increase during the pandemic was proportionate to the increase in all-cause mortality rates. We further calculated mortality rates during pandemic years as compared to three preceding prepandemic years for DPM, other autoimmune diseases with the risk of immune suppression and infection including myasthenia gravis and multiple sclerosis, and diseases with little or no concern for immune suppression (including muscular dystrophy and thyroid gland disorders). As shown in Supplemental Table S3, mortality rates increased by 15% during the pandemic for DPM as compared to 28% for myasthenia gravis, 20% for multiple sclerosis, 7% for muscular dystrophy, and 29% for thyroid gland disorders. Consistently, a retrospective analysis of the TriNetX database between January 2020 and August 2021 showed that individuals with dermatomyositis with COVID-19 had a lower risk of death in comparison to the general population with COVID-19.[40] Finally, the proportions of deaths attributed to COVID-19 as the underlying cause of death during the pandemic were different in three autoimmune disorders, ranging from 8% in multiple sclerosis to 10% in DPM and 14% in myasthenia gravis (Supplemental Table S4).

Limitations of the study using mortality databases include the lack of individualized data on clinical, laboratory, and treatment variables, as well as difficulty in ascertaining the accuracy of the physicians' coding on death certificates.[14,15,41] However, misclassification of an autoimmune disease (ie, recording an autoimmune disease as the cause of death on death certificates for decedents who did not have that autoimmune disease) is unlikely to have influenced our inferences in a substantial way.[15] Unlike administrative health care databases, where possible and probable diagnoses are often recorded, death certificates usually have the most proximate illness recorded as the cause of death. Consequently, DPM is unlikely to be recorded as a cause of death for decedents who did not have DPM. However, DPM may be underrecorded as the cause of death in some proportions of patients whose proximate causes of death were cancer, infection, and ischemic heart disease,[5,6] as suggested by a previous study that showed that DPM was recorded as the underlying (primary) cause of death in only about half of all DPM deaths and as a contributing cause of death in the other half.[12] Therefore, we used the MCD database to capture all DPM deaths across the United States in our study. A previous study showed that the survival rates were similar for dermatomyositis and polymyositis.[7] Hence, we analyzed all DPM deaths together instead of separately analyzing mortality for individual DPM subsets, which might differ. Finally, inclusion of data from 2020 to 2022 allowed us to assess the impact of the COVID-19 pandemic on DPM mortality. However, data for the years 2021 and 2022 were retrieved from the Provisional Mortality Statistics database, which may undergo minor updates when finalized.

Nevertheless, our study provides an unbiased assessment of population-based burden of annual DPM mortality using a large

sample size. Such mortality statistics from national datasets serve as important indicators of the disease-specific burden at the population level, with implications for health policy planning and resource allocation. Use of direct age standardization to calculate annual mortality rates and use of joinpoint regression as a computational approach to identify long-term mortality trends are other strengths of our approach. Finally, computation of the ratio of DPM to all-cause (non-DPM) mortality to assess DPM mortality in the context of overall changes in mortality in the general population is another unique strength of this study.

In conclusion, we report a steady decrease in DPM mortality relative to all-cause mortality over the last two decades. It is also encouraging that there was no disproportionate increase in DPM-specific mortality during the COVID-19 pandemic. However, significantly higher odds of premature death from DPM are concerning, highlighting the need to identify its causes, improve screening to facilitate early diagnosis, and earlier intervention. Studies are needed to determine if more aggressive screening and treatments at early stages of active myositis and interstitial lung disease will reduce the risk of premature death in DPM. A persistently higher DPM mortality rate in non-Hispanic Black individuals also warrants targeted efforts to identify high-risk subpopulations and evaluate socioenvironmental and biologic determinants of adverse DPM outcomes in these populations. Finally, large-scale, population-based, prospective studies are needed to verify differences in DPM mortality by race and ethnicity and age and understand causes underlying such differences, which, along with targeted public health resource allocations, may reduce disparities in DPM outcomes.

ACKNOWLEDGMENTS

We sincerely thank Mr Snehin Rajkumar for assistance with the database, Dr Sangmee Sharon Bae for suggestions on treatment advances in myositis, and Dr Eric Yen for guidance with making trend line figures.

AUTHOR CONTRIBUTIONS

Both 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 Singh 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. Singh RR, Malaviya AN. Dermato-polymyositis: a recent outbreak. J Assoc Physicians India 1988;36(11):678–679.
2. Connolly CM, Gupta L, Fujimoto M, et al. Idiopathic inflammatory myopathies: current insights and future frontiers. Lancet Rheumatol 2024;6(2):e115–e127.
3. Lundberg IE, Fujimoto M, Vencovsky J, et al. Idiopathic inflammatory myopathies. Nat Rev Dis Primers 2021;7(1):86.
4. Prasad ML, Sarkar C, Roy S, et al. Idiopathic inflammatory myopathy: clinicopathological observations in the Indian population. Br J Rheumatol 1992;31(12):835–839.
5. Lundberg IE, Forbess CJ. Mortality in idiopathic inflammatory myopathies. Clin Exp Rheumatol 2008;26(5 suppl 51):S109–S114.
6. Marie I. Morbidity and mortality in adult polymyositis and dermatomyositis. Curr Rheumatol Rep 2012;14(3):275–285.
7. Schiopu E, Phillips K, MacDonald PM, et al. Predictors of survival in a cohort of patients with polymyositis and dermatomyositis: effect of corticosteroids, methotrexate and azathioprine. Arthritis Res Ther 2012;14(1):R22.
8. Danieli MG, Gambini S, Pettinari L, et al. Impact of treatment on survival in polymyositis and dermatomyositis. A single-centre long-term follow-up study. Autoimmun Rev 2014;13(10):1048–1054.
9. Aggarwal R, Charles-Schoeman C, Schessl J, et al; ProDERM Trial Group. Trial of intravenous immune globulin in dermatomyositis. N Engl J Med 2022;387(14):1264–1278.
10. Johnson C, Pinal-Fernandez I, Parikh R, et al. Assessment of mortality in autoimmune myositis with and without associated interstitial lung disease. Lung 2016;194(5):733–737.
11. Hochberg MC, Lopez-Acuna D, Gittelsohn AM. Mortality from polymyositis and dermatomyositis in the United States, 1968-1978. Arthritis Rheum 1983;26(12):1465–1471.
12. Qiao P, Guo Q, Gao J, et al. Long-term secular trends in dermatomyositis and polymyositis mortality in the USA from 1981 to 2020 according to underlying and multiple cause of death mortality data. Arthritis Res Ther 2023;25(1):16.
13. Kronzer VL, Kimbrough BA, Crowson CS, et al. Incidence, prevalence, and mortality of dermatomyositis: a population-based cohort study. Arthritis Care Res (Hoboken) 2023;75(2):348–355.
14. Yen EY, Singh RR. Brief report: lupus-an unrecognized leading cause of death in young females: a population-based study using nationwide death certificates, 2000-2015. Arthritis Rheumatol 2018;70(8):1251–1255.
15. Yen EY, Singh DR, Singh RR. Trends in systemic sclerosis mortality over forty-eight years, 1968-2015: a US population-based study. Arthritis Care Res (Hoboken) 2021;73(10):1502–1510.
16. Yen EY, Shaheen M, Woo JMP, et al. 46-year trends in systemic lupus erythematosus mortality in the United States, 1968 to 2013: a nationwide population-based study. Ann Intern Med 2017;167(11):777–785.
17. Yen EY, Singh DR, Singh RR. Steady decrease in systemic sclerosis mortality rates at younger ages over the past five decades. Rheumatology (Oxford) 2024;63(2):466–471.
18. 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.
19. Li L, D'Silva KM, Lu N, et al. Mortality trends in polymyositis and dermatomyositis: a general population-based study. Semin Arthritis Rheum 2020;50(5):834–839.
20. D'Silva KM, Li L, Lu N, et al. Persistent premature mortality gap in dermatomyositis and polymyositis: a United Kingdom general population-based cohort study. Rheumatology (Oxford) 2021;60(6):2653–2660.
21. Chung MP, Paik JJ. Past, present, and future in dermatomyositis therapeutics. Curr Treatm Opt Rheumatol 2022;8(4):71–90.
22. Dalakas MC, Hohlfeld R. Polymyositis and dermatomyositis. Lancet 2003;362(9388):971–982.

23. Jablonski R, Bhorade S, Strek ME, et al. Recognition and management of myositis-associated rapidly progressive interstitial lung disease. Chest 2020;158(1):252–263.
24. Fujisawa T, Hozumi H, Kono M, et al. Prognostic factors for myositis-associated interstitial lung disease. PLoS One 2014;9(6):e98824.
25. Damoiseaux J, Vulsteke JB, Tseng CW, et al. Autoantibodies in idiopathic inflammatory myopathies: clinical associations and laboratory evaluation by mono- and multispecific immunoassays. Autoimmun Rev 2019;18(3):293–305.
26. Rider LG, Aggarwal R, Machado PM, et al. Update on outcome assessment in myositis. Nat Rev Rheumatol 2018;14(5):303–318.
27. Dalakas MC, Illa I, Dambrosia JM, et al. A controlled trial of high-dose intravenous immune globulin infusions as treatment for dermatomyositis. N Engl J Med 1993;329(27):1993–2000.
28. Yokoyama Y, Furuta S, Ikeda K, et al. Corticosteroid-sparing effect of tacrolimus in the initial treatment of dermatomyositis and polymyositis. Mod Rheumatol 2015;25(6):888–892.
29. Sharma N, Putman MS, Vij R, et al. Myositis-associated interstitial lung disease: predictors of failure of conventional treatment and response to tacrolimus in a US cohort. J Rheumatol 2017;44(11):1612–1618.
30. Oddis CV, Reed AM, Aggarwal R, et al; RIM Study Group. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. Arthritis Rheum 2013;65(2):314–324.
31. Qushmaq KA, Chalmers A, Esdaile JM. Cyclosporin A in the treatment of refractory adult polymyositis/dermatomyositis: population based experience in 6 patients and literature review. J Rheumatol 2000; 27(12):2855–2859.
32. Beckett M, Tan J, Bonnardeaux E, et al. Tofacitinib therapy in refractory inflammatory myositis: a retrospective cohort study of 41 patients. Rheumatology (Oxford) 2024;63(5):1432–1436.
33. Bernatsky S, Joseph L, Pineau CA, et al. Estimating the prevalence of polymyositis and dermatomyositis from administrative data: age, sex and regional differences. Ann Rheum Dis 2009;68(7):1192–1196.
34. Phillippi K, Hoeltzel M, Byun Robinson A, et al; Childhood Arthritis and Rheumatology Research Alliance (CARRA) Legacy Registry Investigators. Race, income, and disease outcomes in juvenile dermatomyositis. J Pediatr 2017;184:38–44.e1.
35. Namsrai T, Parkinson A, Chalmers A, et al. Diagnostic delay of myositis: an integrated systematic review. Orphanet J Rare Dis 2022; 17(1):420.
36. Weisleder H, Valle A, Xie X, et al. Racial disparities in diagnosis and treatment of patients with dermatomyositis of different skin tones. J Clin Rheumatol 2024;30(1):8–11.
37. Nisar MK, Mirza W, Rafiq A, et al. Adherence to biologic therapy - does it vary with ethnicity? Biologicals 2018;54:28–32.
38. Goldstick JE, Cunningham RM, Carter PM. Current causes of death in children and adolescents in the United States. N Engl J Med 2022; 386(20):1955–1956.
39. Mehta B, Jannat-Khah D, Mancuso CA, et al. Geographical variations in COVID-19 perceptions and patient management: a national survey of rheumatologists. Semin Arthritis Rheum 2020;50(5):1049–1054.
40. Pakhchanian H, Khan H, Raiker R, et al. COVID-19 outcomes in patients with Dermatomyositis: a registry-based cohort analysis. Semin Arthritis Rheum 2022;56:152034.
41. Singh RR, Singh DR, Yen EY. Worsening premature death burden gap from systemic sclerosis in men and Black persons: a US nationwide population-based study. J Scleroderma Relat Disord 2023; 8(1):20–26.
42. Lundberg IE, Tjärnlund A, Bottai M, et al. European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Arthritis and Rheumatology 2017; 69(12): 2271–2282. https://doi.org/10.1002/art.40320.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 197–207
DOI 10.1002/acr.25613

Development of a Personalized Visualization and Analysis Tool to Improve Clinical Care in Complex Multisystem Diseases With Application to Scleroderma

Ji Soo Kim,[1] John Scott,[1] Lauren Fisher,[1] Lauren N. Smith,[2] Willie Stewart,[1] Adrianne Woods,[1] Rob Smithwright,[1] Diane Koher,[1] Parastoo Aslanbeik,[1] Aalok B. Shah,[1] Brad Tibbils,[1] Samantha I. Pitts,[1] Ayse P. Gurses,[1] Yushi Yang,[1] Ana-Maria Orbai,[1] Antony Rosen,[1] Laura K. Hummers,[1] Scott L. Zeger,[1] and Ami A. Shah[1]

Objective. In complex diseases, it is challenging to assess a patient's disease state, trajectory, treatment exposures, and risk of multiple outcomes simultaneously, efficiently, and at the point of care.

Methods. We developed an interactive patient-level data visualization and analysis tool (VAT) that automates illustration of the trajectory of a patient with scleroderma across multiple organs and illustrates this relative to a reference population, including patient subgroups who share risk factors with the index patient, to improve estimation of disease state. We conducted VAT usability testing with patients and clinicians. We then embedded results from internally cross-validated, Bayesian multivariate mixed models that calculate an individual's risk of critical events, using baseline risk factors, patient-level information in past trajectories in multiple dimensions, and known outcomes from the entire population and relevant subgroups.

Results. The web-based application aggregates complex, longitudinal data to illustrate patient-, subgroup-, and population-level health trajectories across multiple organ systems. Patients (n = 7) exposed to the VAT reported increased knowledge about their disease and confidence in medical decision-making. Rheumatologists (n = 4) were able to access 8.6 times more data in 81.5% of the time using two-thirds fewer clicks using the VAT compared with an electronic medical record system. Statistical modeling was successfully embedded in the VAT, enabling real-time estimation of a patient's risks of multiple complications.

Conclusion. Systematic analysis and visualization of individual- and population-level data in a complex disease has potential to improve medical decision-making and warrants further study. Individualized risk estimation disseminated at the point of care may enable targeted screening and early intervention in high-risk patients.

INTRODUCTION

In the clinic, physicians taking care of patients with complex rheumatic diseases use cognitive skills to integrate information across multiple parameters and organ systems, factoring in a patient's prior trajectory and baseline risk factors, to make estimates about a patient's health state, risk for complications, and need for high-risk therapies. A key component of this mental process is that it is informed by a physician's prior experiences caring for patients with a similar expression of disease and therefore is not generalizable across providers—particularly in rare diseases or diseases with high clinical variability. Aggregating complex, longitudinal data requires a tremendous time investment on the part of the treating provider who has access to historical and current data only for the patient at hand. It is also challenging to clearly explain this information to patients during

Supported by the Jerome L. Greene Foundation, Johns Hopkins inHealth initiative, the Scleroderma Research Foundation, the Nancy and Joachim Bechtle Precision Medicine Fund for Scleroderma, the Manugian Family Scholar, the Donald B. and Dorothy L. Stabler Foundation, the Chresanthe Staurulakis Memorial Fund, the Sara and Alex Othon Research Fund, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grants P30-AR-070254, R01-AR-073208, and K24-AR-080217).

[1]Ji Soo Kim, PhD, John Scott, BA, Lauren Fisher, MS, Willie Stewart, MA, Adrianne Woods, BD, Rob Smithwright, Diane Koher, MS, Parastoo Aslanbeik, MFA, Aalok B. Shah, M.Eng, Brad Tibbils, BS, Samantha I. Pitts, MD, Ayse P. Gurses, MS, PhD, Yushi Yang, PhD, Ana-Maria Orbai, MD, MHS, Antony Rosen, MBCHB, MS, Laura K. Hummers, MD, ScM, Scott L. Zeger, PhD, Ami A. Shah, MD, MHS: Johns Hopkins University, Baltimore, Maryland; [2]Lauren N. Smith, MD: Georgetown University Medical Center, Washington, District of Columbia.

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.25613).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25613.

Address correspondence via email to Ji Soo Kim, PhD, at jkim478@jhu.edu; or to Ami A. Shah, MD MHS, at Ami.Shah@jhmi.edu.

Submitted for publication September 27, 2024; accepted in revised form July 7, 2025.

SIGNIFICANCE & INNOVATIONS

- In complex, multisystem diseases like scleroderma, there is often significant heterogeneity in patients' clinical manifestations, disease course, and risk of complications. There is a major unmet need to develop tools that aggregate patient-level, longitudinal data and enable continuous learning across populations of patients to improve individualized decision-making.
- We developed an interactive patient-level data visualization and analysis tool that automates illustration of the trajectory of a patient with scleroderma across multiple organs relative to a reference population and calculates real-time estimates of critical events.
- Patients exposed to historical data in the tool reported increased knowledge about their disease and felt better equipped to participate in shared decision-making. Rheumatologists accessed >8 times more data in ~80% of the time compared with the standard electronic medical record and reported higher levels of confidence in their medical decision-making.
- This tool represents a shift in health informatics from a clinician having access to the current patient's data only to contextualizing that patient's data with the experience of a selected subset of prior patients.

routine clinical visits to facilitate shared decision-making. In this study, we sought to develop methods to address these challenges using a prototypic multisystem rheumatic disease, systemic sclerosis (scleroderma).

Scleroderma is a complex disease with high variability in clinical phenotype, longitudinal trajectory, treatment response, and mortality. Scleroderma can affect multiple organ systems, and most organ-specific complications occur in ~15% of patients.[1,2] Several clinically available measures, such as autoantibody type and cutaneous subtype, may be useful to subgroup and risk-stratify patients. However, these population-level risk factors are not easily translatable to clinical practice at the patient level to inform targeted screening or early intervention.[3–5] For example, patients with African ancestry, diffuse scleroderma, or anti–topoisomerase-1 antibodies have a higher risk of clinically significant interstitial lung disease (ILD).[6] However, for an individual patient, it remains unknown how this risk is modified by the presence of multiple risk factors, is affected by her own pulmonary function trajectory early in the disease, or changes with involvement of other organ systems.

To address these challenges, we designed a visualization and analysis tool (VAT) to effectively communicate a patient's trends across multiple organ systems. We developed interactive filters that enable a provider to compare an individual patient to a subgroup of patients who share relevant clinical and biologic characteristics. We then implemented this prototype in a web-based application (called Patient Insight) that can be viewed within different electronic medical record (EMR) systems to bring the tool within clinicians' workflow and enable future dissemination.[7] We conducted a study to evaluate effectiveness and usability of Patient Insight for patients and clinicians. Lastly, using Bayesian multivariate hierarchical models and a cross-validated sequential prediction (CVSP) algorithm, we compute and display real-time personalized risk estimates for major complications, harnessing knowledge from a patient's prior trajectory and known outcomes from patients with similar subgroup characteristics.[8] In this article, we present these design and implementation methods. Although the initial development has focused on scleroderma, we anticipate these methods, employing clinical, computational, and engineering skills, will have broad applicability across complex, multisystem diseases and health systems.

METHODS

Development of the VAT. A key objective of a VAT is to make quantitative the qualitative steps that clinicians currently use to evaluate their prior experience and knowledge when reaching a clinical decision for an individual patient. Supplemental Figure 1 describes a pipeline and series of clinical, statistical, and engineering methods that were used to develop the VAT and embed it within the clinical workflow. Expert clinicians selected clinical parameters, calculations, and interactive features to be displayed. Based on this initial sketch, statisticians developed a flexible prototype R Shiny App that visualizes patient- and population-level longitudinal data, enabling quick iterations and revisions.[9,10] After clinician approval, the Johns Hopkins Medicine (JHM) Technology Innovation Center implemented R Shiny App features into a web-based application that could be viewed within the EMR, which for our institution is Epic. This step met two objectives: (1) to generate a version of the tool physicians can use directly to test its clinical value and (2) to enable future dissemination of the tool across health systems and EMR platforms. The web-based version, referred to as Patient Insight, can be updated outside of the EMR, allowing for rapid improvements to incorporate feedback.

Data sources. This study used clinical data from patients who consented to participate in the Johns Hopkins Scleroderma Center Research Registry, a dynamic-entry, prospective longitudinal cohort. Data from registry participants are ingested into the JHM Precision Medicine Analytics Platform (PMAP) (Supplemental Figure 2), in which multiple data sources are harmonized. JHM's Epic EMR captures all transactions comprising clinical care: clinical visits, laboratory measurements, prescriptions, imaging, and procedures. We use real-time Epic data from PMAP as input to our VAT.

Individualized clinical data visualization and comparison to user-defined subgroups. Patient Insight illustrates a patient's aggregate clinical phenotype in a snapshot view; for scleroderma, this includes cutaneous subtype (ie, diffuse versus limited),[11] disease manifestations, onset dates, and autoantibody status. Longitudinal data are illustrated across multiple organ systems (eg, cutaneous, cardiac, pulmonary, gastrointestinal, renal, peripheral vasculature, and musculoskeletal), including patient-reported outcomes and laboratory data. Longitudinal immunosuppressive medication exposure is shown to assess whether drug exposure alters trajectory in these parameters. Relevant comorbid conditions, such as coronary artery disease, diabetes, hypertension, and cancer, are captured.

We defined critical events by either having longitudinal observations exceed or fall below prespecified thresholds or having a discrete event occur at a particular date. These events were plotted on a single time scale starting from scleroderma onset.[12] Details regarding chosen variables are provided in the Supplemental Materials.

The tool illustrates the 10th, 50th, and 90th percentiles for the entire scleroderma population as a reference group. By plotting individuals' trajectories on top of these reference lines, we can visualize a patient's disease course compared with others. Moreover, we programmed filters to compare a patient's trajectory to a user-specified subgroup based on demographic, clinical, and biologic characteristics. This allows clinicians to easily monitor a patient's disease course relative to a group of similar patients based upon known risk factors, such as age, race, sex, cutaneous subtype, and autoantibody status.

Testing effectiveness and usability of Patient Insight. We conducted two studies to assess patients and rheumatologists' perspectives of using Patient Insight. We sought to gain early feedback on the functionalities of visualizing historical clinical data and comparing individuals' trajectory to a user-defined subgroup. Of note, this component of the study did not include incorporation of personalized predictions of the future risk of critical events, which is described further below. This study was approved by the JHM Institutional Review Board (IRB000252437). Survey questions are in the Supplemental Materials.

Patient perspectives. We recruited seven patients with scleroderma who had at least one year of clinical data and were able to participate in study visits over Zoom during the COVID-19 pandemic. Before exposure to Patient Insight, patients completed a survey to capture their understanding of their disease and its trajectory. Then, patients were introduced to their own data in Patient Insight and shown how providers would navigate the tool and share data during a clinical visit. Organ-specific trends were illustrated relative to a population of >4,000 patients with scleroderma and relative to the subgroup of patients who shared key characteristics. Trends were explained to patients in relation to immunosuppressive therapy exposures when applicable. We conducted semistructured interviews (see Supplemental Materials) to evaluate patients' impressions when the tool was part of the communication process. The same survey questions were administered after the interview to assess if the tool changed participants' understanding of their scleroderma.

Interviews were audio-recorded and transcribed verbatim. Two investigators developed an initial coding structure and identified general themes based on the interview guide and first three interviews.[13,14] The coding structure was enriched in breadth and depth based on subsequent interviews and emergent themes. After achieving thematic saturation with the fifth interview, we conducted two additional interviews to ensure no additional themes were identified. Participant comments were coded and analyzed for sentiment to summarize themes and illustrate key findings.

Provider perspectives. To assess rheumatologists' perspectives, Patient Insight was compared with standard of care using Epic. Testing was performed in the context of four real patient cases. Two cases each were selected to reflect common issues arising in clinic: (1) whether immunosuppression needs to be initiated or changed to treat one or more organ systems and (2) whether additional, potentially invasive testing is needed to assess for complications. For each scenario, academic rheumatologists who regularly see patients with scleroderma were asked to assess, as if they were reviewing data for a clinic visit, one case using Epic and one case using Patient Insight in randomized order. Task duration, number of clicks, and data points reviewed were recorded for each task and analyzed using a paired *t*-test.

Rheumatologists filled out an additional questionnaire to assess: whether Patient Insight helped them understand a patient's health state and trajectory; whether it would help them explain health state and trajectory to patients; and whether they gained any new insight into a patient's disease by seeing data visually. Throughout, a human-factors engineer solicited feedback to understand potential areas for improvement, clarity of data elements, and user perception of task ease or difficulty. Participants were asked to elaborate on whether Patient Insight would change their clinical practice, specifically (1) the time required to prepare for seeing a patient, (2) if they would plan to review the tool with patients when discussing their clinical impression, and (3) which parameters they would most likely share with patients. Following each interview, rheumatologists were emailed a System Usability Scale (SUS) to objectively measure their satisfaction with Patient Insight.[15,16] SUS scores range from 0 (worst) to 100 (best), with 68 considered the threshold of acceptable usability.

Estimation of individuals' health trajectories and prediction of future risk of critical events. A patient's estimated disease trajectories foretell the risk of having extreme

values in the near future. By fully using information in multiple longitudinal measures, we can maximize the precision of estimates of patient-specific and population trajectories. We modeled the longitudinal latent health state using a Bayesian multivariate linear mixed-effects model[17,18] for cardiopulmonary parameters (percent predicted forced vital capacity [pFVC], percent predicted diffusing capacity for carbon monoxide [pDLCO], left ventricular ejection fraction [LVEF], and right ventricular systolic pressure [RVSP]). Details of the model are in the Supplemental Materials. Using model estimates, we calculate and display disease trajectories for each patient for pFVC, RVSP, and EF, which are important surrogates for key outcomes (ILD, pulmonary hypertension [PH], and cardiomyopathy).

Observations falling below or rising above a clinically set threshold are useful surrogates for critical events that will likely require immediate medical attention sometimes followed by more invasive and higher risk interventions. To predict critical events, we projected each individual's trajectory into the future and then calculated the probability the person will cross the following boundaries: LVEF < 50% and LVEF < 35% (cardiomyopathy), RVSP ≥ 45 mmHg and RVSP ≥ 50 mmHg (PH), and pFVC ≤ 70% and pFVC ≤ 60% (ILD) in the next 6, 12, 18, and 24 months. We refer to the estimation of these event probabilities using our methodology as CVSP.[8,19] The CVSP algorithm sequentially produces the risks of events as additional data are observed in real-time for a patient. The predictions are made without refitting the model to incorporate new observations for a patient using a cross-validation method, yielding computational efficiency. We previously showed CVSP increases in precision as more data are observed for a given patient and that, even with no observations for an individual's measure, yields predictions with considerable precision compared with other methods, including random forest classification and logistic regression.[8,19]

We added the cross-validated statistical models that generate patient-specific risk estimates of critical events to the prototype VAT. The feature is now being added to Patient Insight for external validation and calibration.

RESULTS

This study used clinical data of >4,000 patients with scleroderma. Details regarding cohort characteristics are provided in the Supplemental Materials. Figure 1 is a snapshot of Patient Insight currently used by physicians in the clinic to effectively assess patients' past and current health status. A full view of the tool is provided in Supplemental Figure 3. In a single-page view, the VAT displays an individual patient's longitudinal immunosuppression exposures, critical events, trajectory across multiple organ systems, disability index, and comorbidities. Users can select organ systems or variables of interest to only view selected data if desired.

Trajectory relative to a reference population. An important feature of the VAT is presenting an individual patient's data relative to others with the same disease and specifically those who share key demographic, clinical, and biologic characteristics. Filters can be deployed to compare an individual patient to others who share risk factors that associate with disease severity in population studies, such as race and cutaneous and autoantibody type. Figure 2 demonstrates the use of this feature. In Figure 2A, the FVC trajectory of a White woman who developed diffuse scleroderma with anti–Scl-70 antibodies at age 40 years is shown relative to the entire population of patients with scleroderma. In Figure 2B, her same trajectory is shown relative to other patients with similar demographic and clinical characteristics. After 10 years of disease, her pulmonary function has been poorer than the average patient with scleroderma but is typical of that expected for patients with the same race, disease subtype, and autoantibody type. The ability to interactively change the reference population provides insight into how individual and combinations of risk factors may modify a patient's likely trajectory and outcome; this approach furthers development of a continuous learning health system.

Results from patient and provider surveys and interviews. *Patient perspectives.* Among seven patients, four were women and six had diffuse scleroderma, and 51 years was the median age. Self-reported race and education level are provided in Supplemental Tables 1 and 2. Patients reported an increased understanding of their disease and potential prognosis after being exposed to Patient Insight (Supplemental Table 3). Figure 3 shows three main themes identified from the interviews with representative quotes. Many patients expressed that visualizing their data relative to similar patients increased their knowledge about their own disease. They shared that the tool improved recall of past symptoms and relayed clinically relevant patterns that were otherwise difficult to describe. They reported that Patient Insight would allow them to participate in their own medical care more actively, both in the context of clinical care and everyday modifiable behaviors. Some patients mentioned that the tool could help them make the most out of short appointments and increase the efficiency of medical visits. In summary, being exposed to Patient Insight satisfied patients' curiosity about their disease state and enabled them to feel empowered dealing with their scleroderma.

Provider perspectives. Four rheumatologists participated in the study. They were in rheumatology practice (including fellowship) for 5 to 46 years, with 6 to 8 years of experience using Epic. Two described their comfort level with technology as "somewhat comfortable" and two as "very comfortable."

The streamlining and display of historical patient data in Patient Insight resulted in more efficient examination of clinical cases (Supplemental Table 4). On average, providers accessed 8.6 times more data ($t = -8.74$; $P < 0.001$) in 81.5% of the time

Figure 1. Patient Insight tool for scleroderma. This web application illustrates an individual patient's demographic and clinical disease manifestations and trajectory across multiple organ systems. A partial view with selected organ systems is shown in the figure. A complete view of the tool with all organ systems (cardiac, pulmonary, cutaneous, gastrointestinal, renal, peripheral vascular, and muscle) is shown in Supplemental Figure 3. The severity of disease state is indicated by red (severe) and pink (mild) regions for left ventricular ejection fraction and RVSP. Users can view values for each point by hovering over the points in the graphs. DLco, diffusing capacity for carbon monoxide; EF, ejection fraction; FVC, forced vital capacity; GLI, Global Lung Function Initiative; ILD, interstitial lung disease; mmf, mycophenolate mofetil; mRSS, modified Rodnan skin score; po, by mouth; RP, Raynaud phenomenon; RVSP, right ventricular systolic pressure; Rx, prescription; TFR, tendon friction rubs.

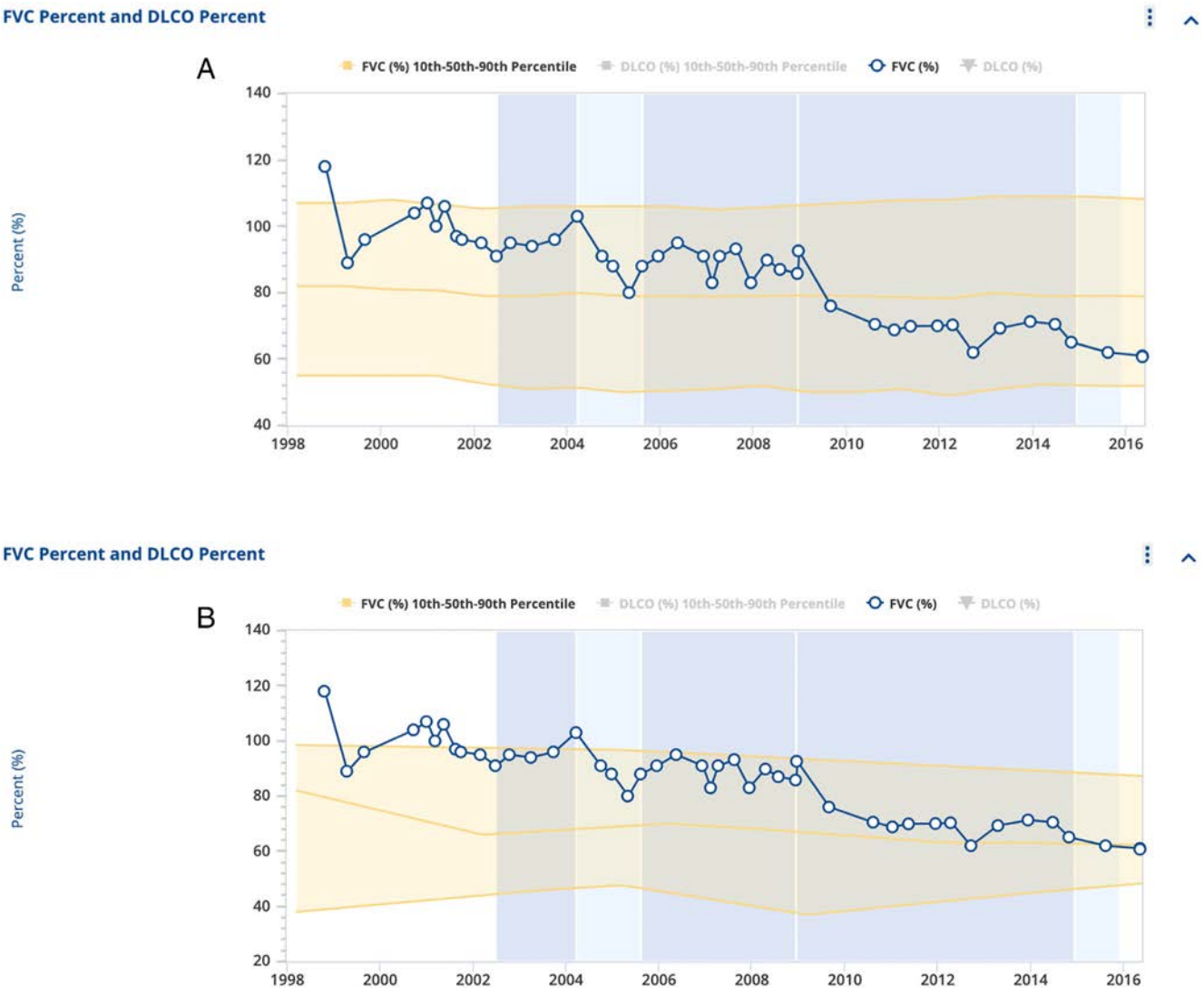


Figure 2. Lung trajectories for an individual patient with interactive filters for the reference population. (A) Lung trajectories of a patient and tenth, median, and ninetieth percentiles reference lines of the overall scleroderma population. (B) Lung trajectories of the same patient and tenth, median, and ninetieth percentiles reference lines of selected subpopulation. On the top panel (A), this patient's trajectory is compared with that of the overall scleroderma center cohort. On the bottom panel (B), that same patient's trajectory is compared with a reference population that is similar to the patient (onset age of 30 to 50 years, diffuse type, and Scl-70–antibody positive). These data illustrate how the reference population changes the interpretation and one's perspective. When comparing this patient to the overall scleroderma population, we observe that her percent predicted FVC trajectory declines from the ninetieth percentile to the tenth over 20 years of follow up, dropping rapidly below the fiftieth percentile line after 10 years. Relative to other similar patients, however, we see that her percent predicted FVC trajectory is better than that typically expected for the first 10 years of follow up and around the median afterwards. DLco, diffusing capacity for carbon monoxide; FVC, forced vital capacity.

(t = 1.76; P = 0.123), using 66.7% fewer clicks (t = 4.17; P = 0.004). During interviews (Figure 3), the rheumatologists reported that Patient Insight helped them assess patients' disease better, with increased efficiency and greater confidence in decision-making. They expressed appreciation for having patient data in one place, highlighting the anticipated time-savings and expressing excitement to share the tool with patients who value having their data presented clearly and concisely. The rheumatologists noted that they gained new insights into a patient's disease and were better able to refine diagnoses using Patient Insight. Those who recommended a change in testing or treatment reported higher levels of confidence in their decision when using Patient Insight versus Epic alone. Providers perceived usability of Patient Insight as "excellent" (SUS scores: mean 81.3 ± 7.5), expressing that the tool was easy to use, and they would like to use it frequently (Supplemental Table 5).

Estimating a patient's risk of critical events. Although visualizing a patient's historical trajectory offers tremendous value, our goal is to improve prediction of future events to make better medical decisions. We developed a prototype that enables us to estimate and visualize each patient's disease trajectory with the

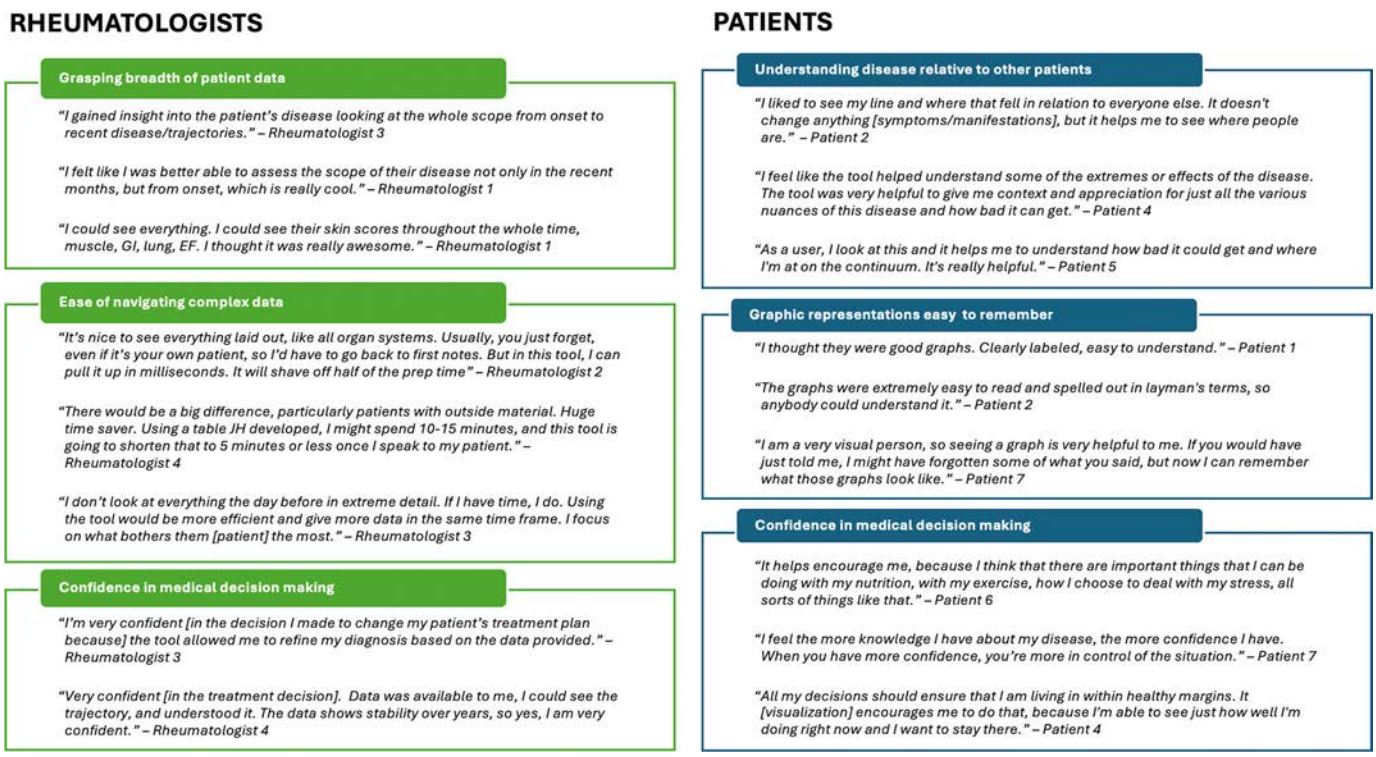


Figure 3. Themes identified from rheumatologist and patient interviews with representative quotes. Rheumatologists reported gaining new insights into a patient's disease, increased efficiency, and greater confidence in decision-making by using the Patient Insight tool. Patients felt that using the tool increased knowledge, improved recall of their past symptoms, and instilled greater confidence in dealing with their disease. EF, ejection fraction; GI, gastrointestinal; JH, Johns Hopkins. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25613/abstract.

95% prediction interval for pFVC, EF, and RVSP by jointly modeling multiple measures (Figure 4). The estimated risks of the three outcomes, each with two severity levels, in the next six months are displayed. We project the estimated trajectory forward and calculate the risks of the future events, shown by the shaded regions in the graphs and tables. The patient in Figure 4 is at high risk of progressive ILD (42% risk for pFVC ≤ 70%), cardiomyopathy (24% for LVEF < 50%; 8% for LVEF < 35%), and PH (29% for RVSP > 50 mmHg; 39% for RVSP ≥ 45 mmHg), suggesting more intensive screening and monitoring is warranted. The predictive analytics are now being incorporated into Patient Insight to test its value in clinical settings.

An individual's estimated health trajectory and risks of critical events are calculated by harnessing information in baseline characteristics (age of scleroderma onset, sex, race, and cutaneous and autoantibody type) and longitudinal cardiopulmonary biomarkers (pFVC, pDL$_{CO}$, LVEF, and RVSP). In Figure 5, we display how using baseline characteristics and subsequently longitudinal biomarkers changes precision in predicted values. When no patient-level information is available, our best estimate for pFVC one year from current visit for pFVC < 70% for very early disease (first 18 months) is depicted in Figure 5A. Probabilities of an individual having a clinically meaningful restrictive ventilatory defect suggestive of ILD is depicted by the areas under the curve. Because no individual-level information is available, the predicted probabilities are derived from the cohort mean, and we observe a widely spread-out bell-shaped curve, reflecting high uncertainty in the predicted value.

We then incorporate an individual's demographic and clinical characteristics to obtain subgroup-specific ILD risk. The estimated risks of ILD for two selected subgroups are shown in Figure 5B: (1) Black women with diffuse scleroderma and anti–Scl-70 (top) and (2) White men with limited scleroderma and anticentromere (bottom). The first subgroup has a higher ILD risk compared with the population-level estimates in Figure 5A. The second subgroup has a lower projected ILD risk.

Individualized estimated ILD risks for two patients with very early disease from each of the subgroups are shown in Figure 5C. The top plot shows patient-specific estimated ILD risk in the next year of an individual patient from the first subgroup. The bottom plot shows ILD risk of a patient from the second subgroup. Having pFVC observations >70% lowered the estimated patient-specific ILD risks compared with subgroup-specific risks for both patients. For the patient in the top plot, the estimated risk of pFVC ≤ 70% decreased from 66% to 15%. For the patient in the bottom plot, this decreased from 15% to 3%. Although the pFVC observations of the two patients are similar with upward trend, the patient in the top plot has a

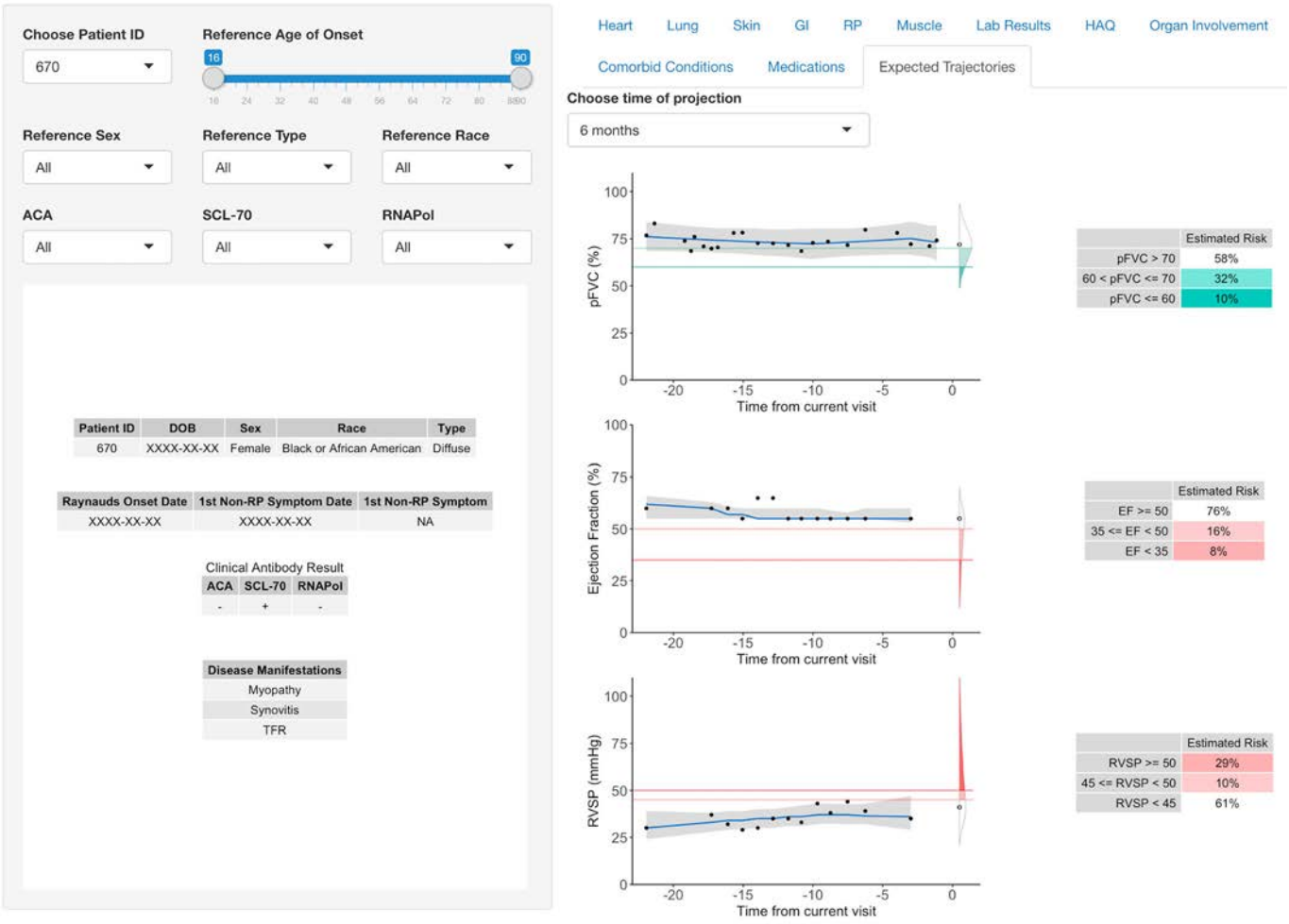


Figure 4. Predictive modeling incorporated in prototype R Shiny App. The figures illustrate an individual patient's trajectory in pFVC, EF, and RVSP over time with the estimated risk of crossing high-risk thresholds in these parameters within the next six months. The degree of uncertainty around these estimates is illustrated in the distribution of potential values projected forward at time 0 (the time of the current visit) and the estimated risks of crossing certain thresholds is shown in the tables on the right (denoted with lighter and darker shades of green and pink). The time of projection can be changed by the end user to illustrate predicted estimates at 6, 12, 18, or 24 months in the future. EF, ejection fraction; GI, gastrointestinal; HAQ, health assessment questionnaire; pFVC, percent predicted forced vital capacity; RP, Raynaud phenomenon; RVSP, right ventricular systolic pressure; TFR, tendon friction rubs.

higher estimated ILD risk compared with the patient in the bottom plot given her baseline characteristics. This illustrates that the estimated trajectories harness information not only from each patient's pFVC observations but also from subgroup-specific characteristics and other longitudinal biomarkers (eg, pDLco). Incorporating patient-specific information results in increased precision of the predicted values shown by narrower bell-shaped curves compared with Figure 5B.

Although the risks of critical events are calculated based on prespecified thresholds, risk estimates using different cutoffs can be readily calculated from our model. For instance, we can estimate an individual's risk of a 5% or 10% decline in pFVC, which could be useful for patients who have higher baseline pFVC or to identify patients who warrant more intensive monitoring, early therapeutic intervention, or consideration for clinical trial enrollment (Supplemental Figure 4).

Harnessing information across measures is particularly useful when some measures have fewer observations compared with others. In our case, patients generally have fewer observations for cardiac measures (EF and RVSP) and richer data for the pulmonary measures (pFVC and pDLco). When there are only sparse data observed for a measure or a patient, we borrow strength from other correlated measures and from the entire cohort to produce more precise estimates for a given parameter.[8] In Figure 6, we demonstrate how the model generates accurate risk estimates for PH using information not only in RVSP trajectory but also in pDLco, a richer measure. Two patients with similar clinical and demographic

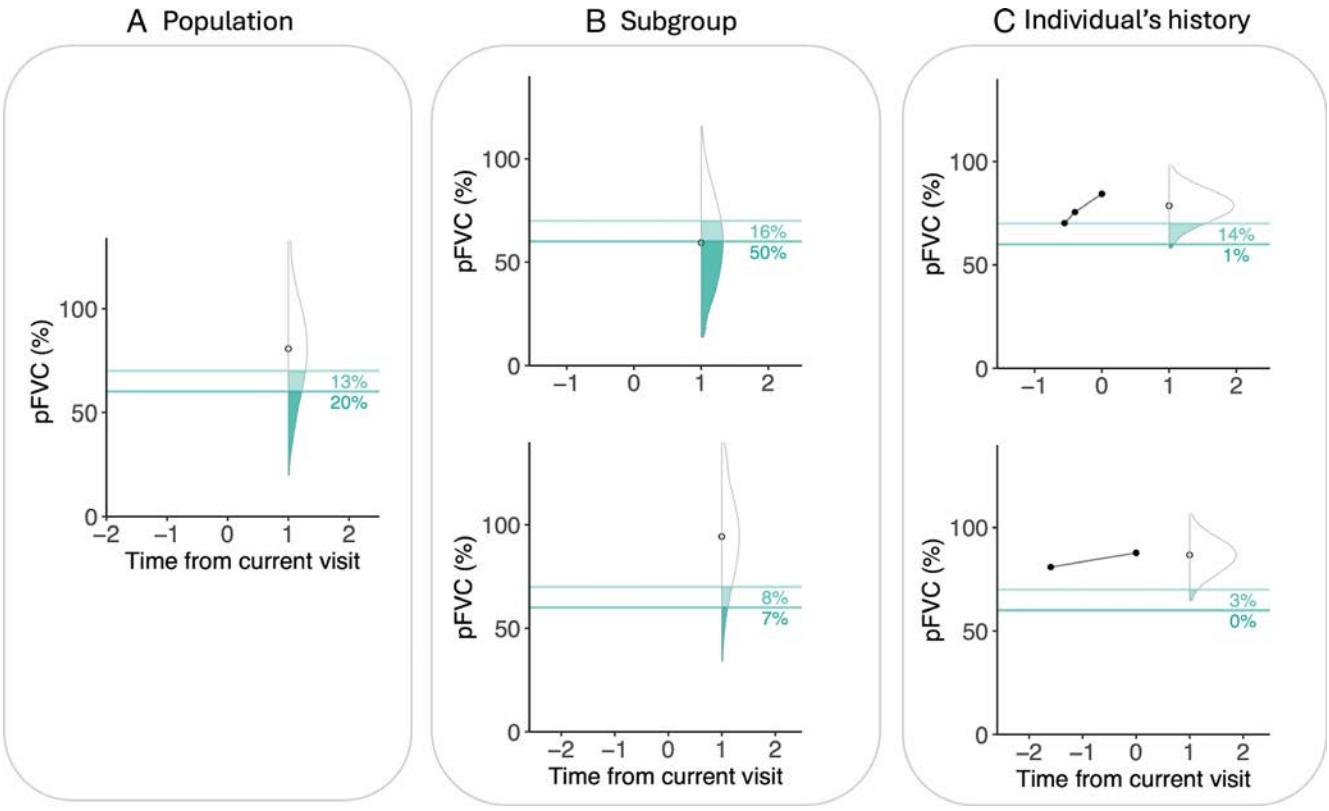


Figure 5. Risks of interstitial lung disease in the overall scleroderma population, within distinct subgroups, and for two individual patients. This series of plots show a gradual reduction in uncertainty for interstitial lung disease prediction as more information is introduced (left to right) for two separate scenarios (top and bottom row). Interstitial lung disease risks are indicated by the shaded regions and percentages. (A) shows expected pFVC distribution for a random patient with scleroderma without any baseline information or follow-up data. The top plot in (B) shows the expected pFVC distribution for a random patient in a selected subgroup: Black women who developed diffuse scleroderma with anti–Scl-70 antibodies. The top plot in (C) shows the interstitial lung disease risk predictions of a patient who has same baseline characteristics with her observed pFVC measurements. In contrast, the bottom plot in (B) shows the expected pFVC distribution for a random patient in a different subgroup: White men who developed limited scleroderma with anticentromere antibodies. The bottom plot in (C) shows the interstitial lung disease risk of a patient who has the same baseline characteristics with his own pFVC observations. Uncertainty in prediction, represented by the width of bell-shaped prediction distributions, decreases when baseline information and subgroup characteristics are known and further decreases substantially when an individual patient's pFVC observations are used. pFVC, percent predicted forced vital capacity.

characteristics are depicted in Figure 6A and 6B, respectively. Both patients are White women with a similar age, skin subtype, and autoantibody status. Both have similar RVSP trajectories up until the time of the current visit, but the estimated risk of PH in the next two years for the patient in Figure 6A is much lower than the patient in Figure 6B. The patient in Figure 6A has an estimated risk of 14% for RVSP ≥ 45 mmHg, whereas the patient in Figure 6B has an estimated risk of 38%. This difference is driven by the patients' pDLCO trajectories. Compared with a similar reference group, the pDLCO trajectory of the patient in Figure 6A is relatively stable across time, fluctuating close to the fiftieth percentile line. However, the pDLCO trajectory in Figure 6B gradually declines over time and falls below the tenth percentile line for the reference group. Indeed, the patient in Figure 6B was subsequently confirmed to have PH by right heart catheterization. This demonstrates how the VAT maximally uses available data to aid a clinician's decision-making.

DISCUSSION

In complex, multisystem diseases like scleroderma, there is often significant heterogeneity in patients' clinical manifestations, disease course, and risk of complications. Clinicians try to incorporate knowledge from a patient's prior trajectory, the interplay across organ systems, and experiences from caring for other similar patients to make an informed estimate of a patient's likely disease course and need for high-risk therapies. In rare or complex diseases with high clinical variability, many providers have insufficient experience to make informed decisions. Further, the challenges of information gathering and processing are substantial, and decision-making may be particularly prone to cognitive biases, such as availability, base rate neglect, and commission biases.[20] There is a major unmet need to develop tools that (1) aggregate patient-level, multisystem data in a format that is easy for clinicians to access, (2) harness knowledge and known

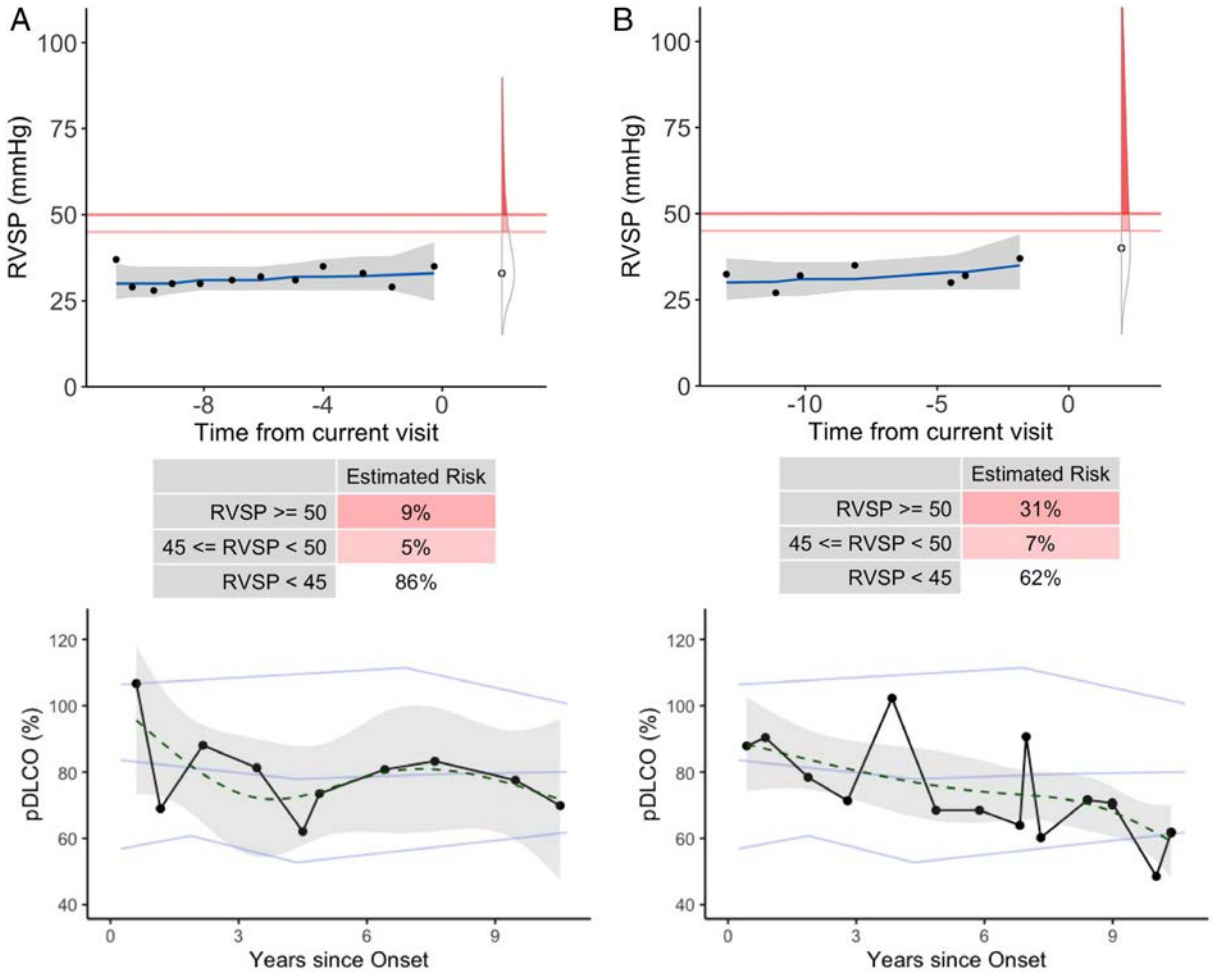


Figure 6. RVSP and pDLco trajectories and estimated risks of pulmonary hypertension of two individual patients. The tables in figures (A) and (B) show estimated risks of crossing high-risk thresholds for PH in the next two years for two patients. Both patients are White women with diffuse cutaneous scleroderma, negative for anti–Scl-70, anti–RNA polymerase III, and anticentromere. Onset ages are 51 and 55 years for patients in (A) and (B), respectively. The reference lines for pDLco trajectories are created using a reference population similar to the patients (White woman, onset age of 40 to 60 years, diffuse type, and Scl-70, RNA polymerase III, anticentromere antibody negative). Smooth curves (green dotted lines) created using each patient's pDLco observations only are shown with their 95% confidence intervals to illustrate the two patients' pDLco trends over time. pDLco, percent predicted diffusing capacity for carbon monoxide; RVSP, right ventricular systolic pressure.

outcomes from other patients who share key characteristics, and (3) compute and display personalized risk estimates to improve decision-making.[21]

This study has several strengths. The novel patient-level VAT shows one patient's trajectory across multiple organs in the context of a clinician-selected, biologically relevant patient subset. The tool also provides precise individualized trajectory and future risk estimates by simultaneously modeling data in multiple organ systems, harnessing data within a patient, across organ systems, and across similar patients. This tool represents a shift in health informatics from a clinician having access to the current patient's data only to contextualizing that patient's data with the experience of a selected subset of prior patients. In our scleroderma example, we confirmed that visualizing historical data in Patient Insight can increase providers' efficiency and enable them to see new trends and patterns in a patient's disease. We also found that the tool improves patients' understanding of their disease. The absence of a cure or common scleroderma trajectory undermined many patients' sense of certainty and control, which was mitigated by Patient Insight.

Limitations of this study include its small sample size, and we are expanding the study to include more rheumatologist and patient participants to test the value of embedded personalized risk estimates to aid in decision-making. It will be important to determine whether the tool enhances a physician's judgment across a range of provider experience levels and whether the tool is interpretable by patients with varying degrees of health literacy. We internally validated the individualized prediction model before embedding it in the prototype, and the predictions' value and usability will be tested after embedding it in the EMR. Further work

is needed to externally validate and calibrate the prediction tools in diverse populations for them to be embedded in EMR systems for wide clinical use.

Development of the VAT required collective effort of clinicians, statisticians, software engineers, user design experts, and human-factors engineers. Developing decision support tools, particularly those that embed predictive analytics, involves serial testing and improvement. We anticipate that the precision of these prediction models can be further improved by harnessing multicenter longitudinal data, adding other outcomes (eg, modified Rodnan skin score trajectory, severe gastrointestinal disease, renal crisis, and myopathy), and incorporating newly discovered molecular or imaging biomarkers. Important goals include modeling the effect of competing medication choices on a patient's likely trajectory and embedding decision support based on predicted probabilities of critical events into the VAT. We envision that sharing this tool with patients during clinical visits will allow patients to better understand their disease and the rationale for whether invasive testing or high-risk medications are needed. Such a tool may improve medication adherence and shared decision-making and reduce decisional regret for patients.

Personalized medicine strategies that harness and integrate knowledge within and across individuals have great potential to improve risk estimation and tailored decision-making in complex diseases. Developing methods to bring new predictive models into clinical settings is critical to demonstrate the value of these tools, foster the creation of a continuous learning health system, and enable future dissemination. The VAT used in this study allows flexible parameterization and can be applied to display clinical data and model health trajectories for other complex diseases and from other data sources. We anticipate this framework will provide a foundation that can be scaled and generalized for multiple clinical applications and disease 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, Drs Kim and A. A. Shah confirm that all authors have provided the final approval of the version to be published and take 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. Muangchan C, Baron M, Pope J; Canadian Scleroderma Research Group. The 15% rule in scleroderma: the frequency of severe organ complications in systemic sclerosis. A systematic review. J Rheumatol 2013;40(9):1545–1556.
2. Shah AA, Hummers LK, Casciola-Rosen L, et al. Examination of autoantibody status and clinical features associated with cancer risk and cancer-associated scleroderma. Arthritis Rheumatol 2015;67(4):1053–1061.
3. Mecoli CA, Shah AA. More than skin deep: bringing precision medicine to systemic sclerosis. Arthritis Rheumatol 2020;72(3):383–385.
4. Igusa T, Hummers LK, Visvanathan K, et al. Autoantibodies and scleroderma phenotype define subgroups at high-risk and low-risk for cancer. Ann Rheum Dis 2018;77(8):1179–1186.
5. Patterson KA, Roberts-Thomson PJ, Lester S, et al. Interpretation of an extended autoantibody profile in a well-characterized Australian systemic sclerosis (scleroderma) cohort using principal components analysis. Arthritis Rheumatol 2015;67(12):3234–3244.
6. Shah AA, Wigley FM. My approach to the treatment of scleroderma. Mayo Clin Proc 2013;88(4):377–393.
7. Scott JW, Patras K, Shah A, et al. A common model for longitudinal visualization of patient data. Poster presented at: Americas Committee for Treatment and Research in Multiple Sclerosis; March 1, 2019; Dallas, TX.
8. Kim JS, Shah AA, Hummers LK, et al. Predicting clinical events using Bayesian multivariate linear mixed models with application to scleroderma. BMC Med Res Methodol 2021;21(1):249.
9. Winston C, Joe C, Allaire JJ, et al. shiny: web application framework for R. 2015.
10. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2020.
11. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15(2):202–205.
12. Butler EA, Baron M, Fogo AB, et al; Scleroderma Clinical Trials Consortium Scleroderma Renal Crisis Working Group. Generation of a core set of items to develop classification criteria for scleroderma renal crisis using consensus methodology. Arthritis Rheumatol 2019;71(6):964–971.
13. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ 2000;320(7227):114–116.
14. Orbai AM, Smith KC, Bartlett SJ, et al. "Stiffness has different meanings, I think, to everyone": examining stiffness from the perspective of people living with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2014;66(11):1662–1672.
15. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, eds. Usability Evaluation in Industry. 1995:189.
16. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. CreateSpace Independent Publishing Platform; 2011.
17. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982;38(4):963–974.
18. Shah A, Laird N, Schoenfeld D. A random-effects model for multiple characteristics with possibly missing data. J Am Stat Assoc 1997;92(438):775–779.
19. Kim JS. Modeling Repeated Multivariate Data to Estimate Individuals' Trajectories, and Risks of Major Clinical Events With Application to Scleroderma. Dissertation. Johns Hopkins University; 2020.
20. O'Sullivan ED, Schofield SJ. Cognitive bias in clinical medicine. J R Coll Phys Edinb 2018;48(3):225–232.
21. Coppola SM, Gurses AP. Remote usability testing of a pediatric trauma dashboard. In: Black NL, Neumann WP, Noy I, eds. Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021). Springer International Publishing; 2021:267–273.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 208–214
DOI 10.1002/acr.25608

Late-Onset Scleroderma Renal Crisis Does Occur—A Multicenter Study

Swati Mehta,[1] Sumbal Wajid,[1] Isam Albaba,[2] Virginia Steen,[3] Robyn T. Domsic-Degazio,[4] Alex Luta,[3] Paul J. Feustel,[5] Loay Salman,[1] and Lee Shapiro[1]

Objective. Scleroderma renal crisis (SRC) is historically described to occur within the first five years of systemic sclerosis (SSc) diagnosis. However, it has been observed to occur beyond five years. In this analysis, we aim to describe the prevalence, clinical features, and outcomes of late-onset SRC and compare them with early-onset SRC.

Methods. This retrospective observational study included patients diagnosed with SRC between 1995 and 2018 at three university-affiliated hospitals. Late-onset SRC was defined as SRC occurring five years after SSc diagnosis. After obtaining institutional review board approval, data including demographics, SRC onset, clinical characteristics, laboratory data, and outcomes were extracted. Continuous data were expressed as mean/median and categorical data as frequencies/percentages. Differences were analyzed via Pearson's chi-square test.

Results. A total of 223 patients with SRC were identified, with 169 (75.8%) classified as early onset and 54 (24.2%) as late onset. The late-onset group had a mean SRC onset of 12.2 years after SSc diagnosis. Male predominance was observed in the late-onset compared with the early-onset group (59% vs 9%). Steroid exposure was more common in late vs early group (56% vs 29%). There was no evidence of difference in autoantibodies profile and outcomes between early- and late-onset groups.

Conclusion. Late-onset SRC was seen in up to 25% of patients with SRC, with a mean of 12 years after SSc diagnosis. Our findings reveal comparable clinical characteristics and outcomes between early- and late-onset SRC, underlying the importance of recognizing SRC regardless of disease duration to optimize outcomes.

INTRODUCTION

Systemic sclerosis (SSc) or scleroderma is an autoimmune connective tissue disease that is characterized by vasculopathy and fibrosis. It is divided into two main subtypes based on the extent and distribution of skin and organ involvement: limited cutaneous systemic sclerosis, which predominantly affects the skin of the face, hands, and fingers, and diffuse cutaneous systemic sclerosis, which is characterized by more widespread skin involvement and an early onset of internal organ manifestations.[1] Scleroderma renal crisis (SRC) is a life-threatening complication of SSc characterized by malignant hypertension and acute kidney injury (AKI).[2] SRC occurs in about 10% of patients with scleroderma.[3] It is more common in the diffuse group, with 20% to 25% of patients in this group developing the complication, compared with only 1% in the limited group.[4] SRC is associated with high morbidity and mortality. With the advent of angiotensin-converting enzyme (ACE) inhibitors, mortality has decreased.[5] Nevertheless, the one-year outcomes for patients with SRC are still concerning, with over 30% experiencing mortality and 25% remaining dependent on dialysis.[6]

The Scleroderma Clinical Trial Consortium identified five domains to develop a classification criteria for SRC: blood pressure, AKI, microangiopathic hemolytic anemia and thrombocytopenia, target organ dysfunction, and renal histopathology.[7] Some of the risk factors associated with the development of SRC include the presence of anti-RNA polymerase III antibodies,[8] the presence of diffuse skin involvement, and antecedent glucocorticoid exposure.[9] Classically, about 80% of SRC occurs within the first five years of diagnosis of SSc.[10,11] However, SRC does occur later in the course of the disease (ie, >5 years of disease).[12–14] The characteristics of this late SRC

Supported by the Scleroderma Clinical Trial Consortium.
[1]Swati Mehta, MBBS, Sumbal Wajid, MBBS, Loay Salman, MD, Lee Shapiro, MD: Albany Medical Center, Albany, New York; [2]Isam Albaba, MD: Temple University, Philadelphia, Pennsylvania; [3]Virginia Steen, MD, Alex Luta, MD: MedStar Georgetown University Hospital, Washington, DC; [4]Robyn T. Domsic-Degazio, MD: University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; [5]Paul J. Feustel, PhD: Albany Medical College, Albany, New York.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25608.
Address correspondence via email to Sumbal Wajid, MBBS, at sumbalwajid8@gmail.com.
Submitted for publication April 30, 2025; accepted in revised form July 2, 2025.

SIGNIFICANCE & INNOVATIONS

- This is the first study, to our knowledge, to systematically characterize the clinical profile of patients with late-onset scleroderma renal crisis (SRC) (SRC occurring beyond five years of systemic sclerosis [SSc] diagnosis) and compare them with early-onset SRC (SRC occurring within five years of SSc diagnosis). The findings fill the gaps in the understanding of SRC presentation across the disease duration.
- Up to 25% of patients developed SRC late in our patient population, with a mean onset of 12 years after the initial diagnosis of SSc. These findings emphasize the need for continued vigilance for SRC regardless of the disease duration to improve timely diagnosis and outcomes.
- Exposure to steroids was significantly associated with late-onset SRC, reinforcing the importance of cautious steroid use. Clinicians should maintain a high index of suspicion and closely monitor for SRC in patients receiving steroids, regardless of disease duration.

onset population have never been studied. Our objective was to describe the prevalence, clinical features, and outcomes of late-onset SRC and compare those with early-onset SRC.

METHODS

Study design and settings. We conducted a retrospective observational study across three university-affiliated hospitals: Albany Medical Center, Georgetown University Hospital, and the University of Pittsburgh. Approval of the study was obtained from the institutional review board at each medical center.

Study population. All adult patients diagnosed with SRC who met the preliminary classification criteria[7] between 1995 and 2018 across the three medical centers were included. The classification criteria incorporated the following: an acute rise in blood pressure, AKI, microangiopathic hemolytic anemia and thrombocytopenia, target organ dysfunction (such as hypertensive retinopathy, hypertensive encephalopathy, acute heart failure, and acute pericarditis), and renal biopsy findings indicative of SRC.

Case identification was performed using pertinent diagnostic codes and chart review at Albany Medical Center and Georgetown University. The following *International Classification of Diseases*, *Ninth Revision* and *Tenth Revision* (*ICD-9* and *ICD-10*) codes were used for the identification of SRC cases: the codes used for scleroderma were *ICD-9* M710.1 and *ICD-10* M34.9/M34.89, which were used in combination with the codes for AKI, *ICD 9* 584.9/593 and *ICD-10* N17/N19. The University of Pittsburgh used their cohort database with SRC cases.

Group classification. SRC was categorized into two groups based on the timing of its onset relative to SSc diagnosis. The groups were early onset (SRC occurring within the first 5 years of SSc diagnosis) and late onset (SRC occurring more than 5 years after SSc diagnosis).

Clinical variables assessed. The clinical features of the late-onset and early-onset SRC groups were compared to evaluate differences in demographics (age, race, gender), the type and duration of scleroderma, autoantibody profile, comorbid conditions, organ system involvement, and outcomes. The comorbidities analyzed included hypertension, Type II diabetes mellitus, chronic kidney disease, heart failure, chronic obstructive pulmonary disease, smoking, overlap syndrome, and exposure to steroids, tacrolimus, and cyclosporine. The overlap syndromes considered were systemic lupus erythematosus, rheumatoid arthritis, dermatomyositis/polymyositis, and Sjögren disease. Organ involvement was assessed across several systems, including the skin, lungs, heart, and gastrointestinal (GI) tract. Skin manifestations included sclerodactyly, finger lesions (ulcerations or scars), and telangiectasias. GI involvement was characterized by oropharyngeal dysphagia, gastroparesis, gastroesophageal reflux disease, gastric antral vascular ectasia, small intestinal bacterial overgrowth, and primary biliary cholangitis. Pulmonary involvement was defined by interstitial lung disease and pulmonary arterial hypertension, whereas cardiac involvement included heart failure, conduction abnormalities, and pericardial involvement. Outcomes evaluated included the need for hemodialysis (HD) at diagnosis and at one year, as well as mortality rates at one and five years.

Statistical analysis. Continuous data are presented as mean ± SD. Differences between two groups were evaluated using the *t*-test for normally distributed data or the Mann–Whitney U test for skewed data. Categorical data are presented as frequencies and percentages. Associations between categorical variables were assessed using Pearson's chi-square test, and results were reported with odds ratio (OR) and 95% confidence intervals (CI). The analyses consisted of univariable associations between clinical variables and timing of SRC, comparing early- and late-onset groups. Statistical analyses were performed using Minitab (version 19) and R (version 4.3.0), with a significance level set at $P < 0.05$.

For the variables having frequency of zero or 100%, ORs were not determined because of quasi-complete separation and failure to fit the logistic model. Missing data were not imputed, and analyses were conducted using all available data.

RESULTS

Time to SRC onset. A total of 223 patients with SRC were identified: 169 (75.8%) were classified with early-onset SRC and

54 (24.2%) with late-onset SRC. The average time from diagnosis of SSc to the time of SRC in early-onset SRC was 0.44 (SD = 1.03) of a year. This included 29 patients who had renal crisis at the time of diagnosis of scleroderma. Patients in the late-onset group had significantly longer disease duration with the mean time from diagnosis of SSc to onset of SRC being 12.2 (SD = 6.5) years (Table 1). Figure 1 illustrates the distribution of time to SRC onset in the late-onset group, revealing that nearly half of the patients in this group (48%) developed SRC more than 10 years after their SSc diagnosis, with five patients who ended up developing it more than 20 years after SSc diagnosis.

Baseline characteristics. There was no evidence of difference in age and race between the two groups. However, a significant male predominance was observed in the late-onset SRC group compared with the early-onset group (59% vs 9%) with 13.9 higher odds (95% CI 6.6–29.4) of developing late-onset SRC in men compared with women (Tables 1 and 2). There was no evidence of difference between the two groups regarding scleroderma phenotype, with more than 80% of both groups having diffuse scleroderma (Tables 1 and 2).

Underlying hypertension was more prevalent in the early-onset group (11%) compared with the late-onset group (2%). However, this was consistent with the frequency of essential hypertension. Similarly, malignancy was more frequently seen in the early-onset group, although there was no evidence of difference. Steroid exposure was more common in the late-onset SRC group (56%) compared with the early-onset group (29%), with the odds of developing late-onset SRC being 3.1 times in patients exposed to steroids compared with patients without steroid exposure (95% CI 1.0–9.6) (Tables 1 and 2).

Autoantibody profile. Autoantibody profiles did not show any difference between early- and late-onset SRC. A total of 56% and 62% of the patients had RNA polymerase III antibodies in the early- and late-onset group, respectively (Table 3).

Clinical presentation. Most of the patients (89%) were hypertensive at the time of diagnosis of SRC. There was no

Table 1. Demographics, comorbidities, risk factors, and clinical characteristics of the participants*

Clinical features	Missing data	Early-onset group, N = 169	Late-onset group, N = 54
Onset of SRC from diagnosis of SSc, mean ± SD, y	–	0.44 ± 1.03	12.2 ± 6.5
Age at SRC, mean ± SD, y	–	52.2 ± 14.7	54.9 ± 11.6
SBP, mean ± SD, mm Hg	–	175 ± 33.9	176 ± 35.3
DBP, mean ± SD, mm Hg	–	100.6 ± 19	101 ± 18.6
Males, n/N (%)	–	16/169 (9)	32/54 (59)
Race, n/N (%)	1/223 (0.4)		
White	–	134/168 (80)	66/54 (85)
Black	–	26/168 (15)	5/54 (9)
Others	–	8/168 (5)	3/54 (6)
Comorbidities, n/N (%)			
Hypertension	–	18/169 (11)	1/54 (2)
Diabetes mellitus	–	10/169 (6)	1/54 (2)
Obesity	67/223 (30)	4/112 (3.5)	0/44 (0)
Malignancy	–	30/169 (18)	4/54 (7.4)
Smoking	5/223 (2.2)	71/165 (43)	29/53 (55)
COPD	2/223 (0.9)	2/168 (1.2)	2/53 (3.8)
Overlap syndrome	–	16/169 (9.5)	9/54 (16.6)
Steroid exposure	142/223 (63.7)	19/65 (29)	9/16 (56)
Scleroderma phenotype, n/N (%)	4/223 (1.8)		
Limited cutaneous	–	31/166 (18.7)	7/53 (13.2)
Diffuse cutaneous	–	134/166 (80.7)	46/53 (86.8)
Sine	–	1/166 (0.6)	0/53 (0.0)
Organ involvement, n/N (%)			
Skin	–	168/169 (99)	54/54 (100)
Raynaud phenomenon	9/223 (4)	141/164 (86)	46/50 (92)
Gastrointestinal, n/N (%)	128/223 (57)	74/86 (86)	21/24 (87.5)
Pulmonary, n/N (%)			
ILD	13/223 (5.8)	59/160 (37)	20/50 (40)
PAH	–	33/169 (20)	10/54 (19)
Cardiac, n/N (%)			
Heart failure	143/223 (64)	23/61 (38)	14/19 (14)
Pericardial	30/223 (13.5)	63/146 (43)	27/47 (57)

* Overlap syndrome includes: scleroderma with Sjögren disease, polymyositis, lupus, and rheumatoid arthritis. COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; ILD, interstitial lung disease; PAH, pulmonary arterial hypertension; SBP, systolic blood pressure; SRC, scleroderma renal crisis; SSc, systemic sclerosis.

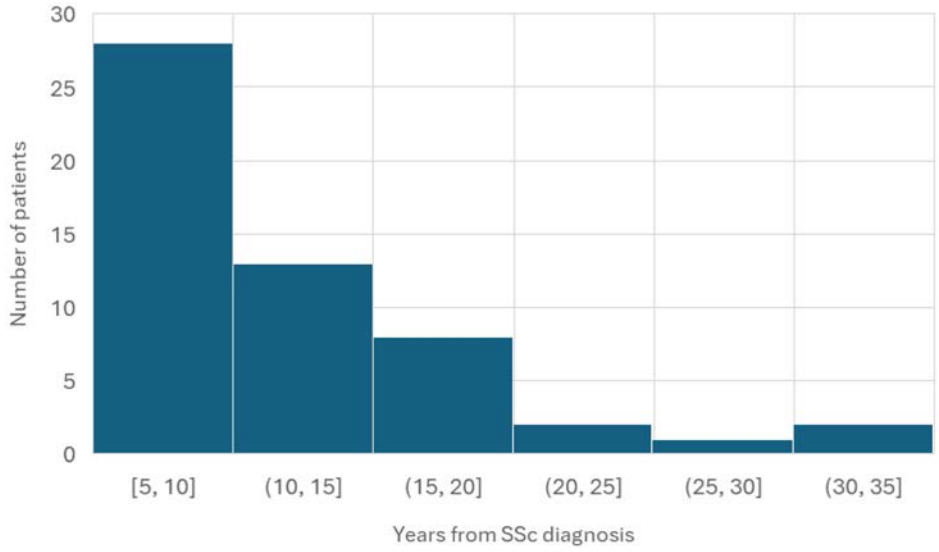


Figure 1. Histogram showing distribution of time from diagnosis of SSc to SRC in late-onset group. SRC, scleroderma renal crisis, SSc, systemic sclerosis.

difference in the level of the first mean blood pressure (BP) at the time of SRC between the two groups. The mean ± SD systolic BP at the onset of SRC was 175 ± 33.9 mm Hg (95% CI 169.9–180.3) for the early-onset group and 176 ± 35.3 mm Hg (95% CI 166.6–186.1) for the late-onset group. The mean ± SD diastolic BP at SRC onset was 100.6 ± 19 mm Hg (95% CI 97.6–103.5) and 101 ± 18.6 mm Hg (95% CI 95.9– 106.1) for the early- and late-onset groups, respectively. A total of 18 patients had normotensive SRC, defined as a BP of 120/80 mm Hg or lower at the time of SRC, with 14 in the early-onset group and 4 in the late-onset group.

There was no difference in the frequency of organ involvement between the two groups except heart failure, which was more common in the early-onset group (38%) compared with the late-onset group (14%), with the OR of 0.2 (95% CI 1.5–14.5) for developing heart failure in late-onset group (Tables 1 and 2).

Table 2. ORs for clinical variables*

Clinical variables	Missing data, N = 223[a]	OR	95% CI
Sex, n (%)			
Reference: female	–	13.9	6.6–29.4
Race	1 (0.4)		
Reference: White	–		
Black	–	0.5	0.2–1.5
Others	–	1.1	0.3–4.3
Comorbidities, n (%)			
Hypertension	–	0.2	0.02–1.2
Diabetes mellitus	–	0.3	0.04–2.4
Obesity[a]	67 (30)	–	–
Malignancy	–	0.4	0.12–1.1
Smoking	5 (2.2)	1.6	0.9–3.0
COPD	2 (0.9)	3.3	0.4–23.7
Overlap syndrome	–	1.9	0.8–4.6
Steroid exposure	142 (63.7)	3.1	1.0–9.6
Scleroderma phenotype, n (%)	4 (1.8)		
Reference: diffuse	–		
Limited scleroderma	–	0.7	0.3–1.6
Organ involvement	–		
Skin[a]	–	–	–
Raynaud phenomenon	9 (4)	1.9	0.6–5.7
Gastrointestinal, n (%)	128 (57)	1.1	0.3–4.4
Pulmonary, n (%)			
ILD	13 (5.8)	1.1	0.6–2.2
PAH	–	0.9	0.4–2.1
Cardiac, n (%)			
Heart failure	143 (64)	0.2	1.5–14.5
Pericardial	30 (13.5)	1.8	0.9–3.5

* CI, confidence interval; COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease; OR, odds ratio; PAH, pulmonary arterial hypertension.
[a] Because of a frequency of an event being zero or 100%, ORs were not determined because of quasi-complete separation and failure to fit the logistic model.

Treatment and outcomes. All patients were treated with ACE inhibitors. There were no differences in renal outcomes between the two groups. The need for HD at the time of SRC diagnosis was 46% in both groups. Of the 77 patients requiring HD at diagnosis in the early-onset group, 39% continued to need it at one year. In comparison, 56% of patients in the late-onset group (14 of 25) still required HD after one year. Long-term outcomes, as indicated by one- and five-year overall mortality rates, did not differ significantly between the groups (Table 4) with more than half of the patients no longer alive at five years.

DISCUSSION

Classically, SRC is thought to occur early in the course of SSc with about 80% of SRC occurring within the first four years of onset of SSc symptoms.[10,15,16] In our study of 223 patients with SRC, late-onset SRC accounted for approximately 25% of

Table 3. Serologies in early- and late-onset SRC*

Serology	Early-onset group, n/N available (%), N = 169	Late-onset group, n/N available (%), N = 54	OR	CIs
Antinuclear antibody[a]	130/131 (99)	44/44 (100)	–	–
Anti-RNA polymerase III antibody	102/163 (62.5)	27/48 (56)	0.8	0.4–1.5
Anticentromere antibody[a]	3/167 (2)	0/51 (0)	–	–
Anti-topoisomerase I antibody	28/161 (17)	13/54 (24)	1.5	0.7–3.2
Anti-U1RNP antibody	5/160 (3)	3/47 (6)	2.1	0.5–9.2
Anti-U3RNP antibody[a]	8/147 (5)	0/43 (0)	–	–
Anti-dsDN	2/150 (1)	0/47 (0)	–	–
Anti-Smith	2/146 (1)	1/47 (2)	1.6	0.1–17.7
Anti-Th/to	3/138 (2)	2/42 (5)	2.3	0.4–13.9
Anti-PMScl	9/143 (6)	3/44 (7)	1.1	0.3–4.2
Anti-Ro (SSA)	10/155 (6)	4/47 (9)	1.4	0.4–4.5
Anti-La (SSB)	7/154 (5)	3/47 (6)	1.4	0.4–5.8

* CIs, confidence intervals; dsDNA, double stranded deoxyribonucleic acid; OR, odds ratio; PMScl, polymyositis/scleroderma, SRC, scleroderma renal crisis.
[a] Because of a frequency of an event being zero or 100%, ORs were not determined because of quasi-complete separation and failure to fit the logistic model.

cases, which is similar to the earlier studies that reported 20% of renal crises occurring four years after the onset of SSc symptoms. However, in our study, it may actually be a bit higher because we defined late onset as five years after the diagnosis of SSc, whereas other studies have used the onset of SSc symptoms, which is typically 6 to 12 months before the diagnosis is made. Thus, our frequency may be even higher if we had chosen four years from the first symptom onset. This emphasizes that there still are a significant number of patients with SRC who develop this complication long after the onset of symptoms. In fact, in the late-onset group, approximately half of the cohort developed it more than 10 years after the diagnosis of SSc. Thus, it is critical for patients and providers to continue to consider the potential of the development of SRC well into the course of the disease and not just early in the disease (Figure 1).

Our most striking finding is the high rate (56%) of recent steroid exposure in the late-onset SRC group. Although we did not have complete data on the dosage of steroids received, this is an important finding. It is possible that both patients and physicians in this group were less vigilant in avoiding steroids, assuming that the patients were not at an increased risk for SRC. This underscores the importance of vigilant steroid use in all patients with SSc, including those in the later stages of the disease. The recent 2023 EULAR recommendations on systemic sclerosis management suggest monitoring patients with SSc for SRC when they are prescribed steroids.[17] Our findings provide further face validity and support for this recommendation.

The late-onset group in our study had a surprisingly prolonged mean time of 12 years from diagnosis of SSc to the diagnosis of SRC. Although we did not collect data on the clinical activity of the SSc itself before the onset of SRC, anecdotally, several of the authors remember that some of these patients with SRC developed SRC in the setting of recurrent or persistent activity of the scleroderma. This may be linked to late flares of scleroderma, a phenomenon that has not been extensively studied. Such flares of the disease in the skin, muscle, joints, or lungs could have led to steroid exposure, which in turn may have increased the risk of developing SRC.[18] In one study, 6% of scleroderma patients had a skin flare of their diffuse cutaneous SSc in a mean of 10 years after having experienced significant improvement in their skin.[19] In 3 of these 39 patients, SRC also occurred in the patients with skin flare. These skin flares were frequently in the setting of patients who had discontinued their immunosuppressive agents. This reduction in immunosuppressive therapy could have triggered skin flares as in the aforementioned study or perhaps renal flares in some of our patients. This hypothesis warrants further investigation; however, because of

Table 4. Outcomes in early- and late-onset SRC*

Patient outcomes	Missing data	Early-onset group, N = 169	Late-onset group, N = 54	*P* value[a]
HD, n/N (%)				
At diagnosis	–	77/169 (46)	25/54 (46)	1.0
At 1 year	–	30/169 (18)	14/54 (26)	0.24
Overall mortality, n/N (%)				
At 1 year	40/223 (18)	28/140 (20)	12/43 (28)	0.27
At 5 years	48/223 (21.5)	70/133 (53)	25/42 (66)	0.43

* HD, hemodialysis; SRC, scleroderma renal crisis.
[a] *P* values sourced through chi-square test.

the retrospective nature of our study, relevant data on this aspect are unavailable.

Interestingly, although there was female predominance in the early-onset group that is consistent with the published data,[20,21] there were significantly more men in the late-onset group. There were no differences in the race or the frequency of anti-RNA polymerase III positivity in the two groups. Impressively, anti-RNA polymerase III was the predominant antibody in 62% and 56% of patients in our early-onset and late-onset SRC groups, respectively. Although a French study[22] reported a lower prevalence of anti-RNA polymerase III, detecting it in only 27% of patients with SRC, our findings did not reflect this trend. Instead, our results aligned more closely with those of Penn et al from England,[15] who also reported a higher prevalence of this antibody in patients with SRC.

SRC was more commonly seen in the diffuse SSc phenotype, both in the early-onset and late-onset groups. Although 31 patients in the early group had limited cutaneous scleroderma, only 3 of them had anti-centromere antibody. Many of these patients had early-stage SSc and likely had not yet developed diffuse cutaneous disease because of the very early duration of their disease.

Early-onset patients with SRC more frequently had hypertension, which can also be seen before and during renal crisis. Severe hypertension was present at the time of SRC in almost 90% of patients, which is typical of what has been seen previously. Although some studies have shown significant increases in previous hypertension in patients with SRC,[23] the frequency of hypertension before SRC in our patients was consistent with what is seen for essential hypertension. Hypertension may be a risk factor for SRC, particularly in Black patients, as the Gordon et al study was from Walter Reed National Military Medical Center in Bethesda, Maryland and included a high frequency of Black patients.[23] The majority of our population, however, was White.

Although malignancy was more frequently observed in the early-onset group, the difference between the two groups was not statistically significant. Nevertheless, there are sporadic case reports describing the occurrence of SRC in individuals with underlying malignancies.[24] This phenomenon has been attributed to a temporal association between the development of anti-RNA polymerase III antibodies and tumor antigen expression.[25] Given that anti-RNA polymerase III antibodies are strongly associated with an increased risk of SRC, this immunologic link may help explain the higher frequency of malignancy observed in patients who develop SRC. There was no clinically significant difference between the early- and late-onset groups in terms of other comorbidities and clinical characteristics. The frequency of organ involvement was comparable between the two groups, with the exception of heart failure at the time of SRC presentation, which was significantly more common in the early-onset group (OR 4.6). Although the odds were lower in the late-onset group at 0.2, this finding may be limited by missing data. These results are consistent with findings from the *American Journal of Emergency Medicine*, which reported that SRC was associated with a significantly increased incidence of new-onset heart failure, with an OR of 4.3 (95% CI 2.6–6.7).[26]

The requirement for HD at the time of renal crisis was similar in both early- and late-onset groups, with almost half of them requiring dialysis. By one year, the overall mortality was also not any different. Otherwise, the clinical, demographic, and renal crisis manifestation were not significantly different in the late-onset SRC compared with early-onset SRC.

Other studies on renal crises have not specifically focused on late-onset renal crises, although many have included cases with late-onset. In the large series by Penn et al,[15] the mean disease duration was only 7.5 months, but they documented cases with onset as late as 16 years. Additionally, a few case reports have described late-onset SRC occurring beyond five years of disease onset.[12–14] These cases often involved steroid exposure, post-renal transplant patients, the use of calcineurin inhibitors, or overlap syndromes with lupus. Our much larger series of renal crisis cases suggests that, although diffuse cutaneous scleroderma and RNA polymerase III antibodies remain prominent risk factors, patients should continue to be closely monitored for late-onset renal crisis, particularly those who have been exposed to steroids.

Our study has several strengths, including a relatively large sample size for a relatively rare disease. Furthermore, the results of the study can be generalized to the general population with SSc, given the multicenter nature of the study. The major limitations of this study include its retrospective design, in which cases were identified through database searches and chart reviews. As a result, there were significant missing data in both groups, which may have impacted our findings. Notably, data on the dosing and timing of steroid exposure before the onset of renal crisis, as well as information on previous treatment with ACE inhibitors/ARBs, were incomplete.

ACKNOWLEDGMENTS

ChatGPT was used to check for spelling and grammatical errors and correction.

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 Wajid 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. Varga J, Abraham D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. J Clin Invest 2007;117(3):557–567.
2. Hudson M, Ghossein C, Steen V. Scleroderma renal crisis. Presse Med 2021;50(1):104063.
3. Steen VD. Scleroderma renal crisis. Rheum Dis Clin North Am 2003; 29(2):315–333.
4. Steen VD. Systemic sclerosis. Rheum Dis Clin North Am 1990;16(3): 641–654.
5. Bose N, Chiesa-Vottero A, Chatterjee S. Scleroderma renal crisis. Semin Arthritis Rheum 2015;44(6):687–694.
6. Hudson M, Baron M, Tatibouet S, et al; International Scleroderma Renal Crisis Study Investigators. Exposure to ACE inhibitors prior to the onset of scleroderma renal crisis-results from the International Scleroderma Renal Crisis Survey. Semin Arthritis Rheum 2014;43(5): 666–672.
7. Butler EA, Baron M, Fogo AB, et al; Scleroderma Clinical Trials Consortium Scleroderma Renal Crisis Working Group. Generation of a core set of items to develop classification criteria for scleroderma renal crisis using consensus methodology. Arthritis Rheumatol 2019;71(6): 964–971.
8. Steen VD. Autoantibodies in systemic sclerosis. Semin Arthritis Rheum 2005;35(1):35–42.
9. Helfrich DJ, Banner B, Steen VD, et al. Normotensive renal failure in systemic sclerosis. Arthritis Rheum 1989;32(9):1128–1134.
10. Steen VD, Medsger TA Jr, Osial TA Jr, et al. Factors predicting development of renal involvement in progressive systemic sclerosis. Am J Med 1984;76(5):779–786.
11. Steen VD. Kidney involvement in systemic sclerosis. Presse Med 2014;43(10 Pt 2):e305–e314.
12. Nunokawa T, Akazawa M, Yokogawa N, et al. Late-onset scleroderma renal crisis induced by tacrolimus and prednisolone: a case report. Am J Ther 2014;21(5):e130–e133.
13. Sanada H, Hara S, Horita M, et al. De novo normotensive scleroderma renal crisis six years after living-donor renal transplantation in a patient with overlapping systemic sclerosis/systemic lupus erythematosus syndrome: a case report. BMC Nephrol 2023;24(1):355.
14. Cheung WY, Gibson IW, Rush D, et al. Late recurrence of scleroderma renal crisis in a renal transplant recipient despite angiotensin II blockade. Am J Kidney Dis 2005;45(5):930–934.
15. Penn H, Howie AJ, Kingdon EJ, et al. Scleroderma renal crisis: patient characteristics and long-term outcomes. QJM 2007;100(8):485–494.
16. Teixeira L, Mahr A, Berezné A, et al. Scleroderma renal crisis, still a life-threatening complication. Ann N Y Acad Sci 2007;1108(1):249–258.
17. 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.
18. Guillevin L, Bérezné A, Seror R, et al. Scleroderma renal crisis: a retrospective multicentre study on 91 patients and 427 controls. Rheumatology (Oxford) 2012;51(3):460–467.
19. Steen V, Medsger TA. Skin flares in systemic sclerosis with diffuse scleroderma (dcSSc) [abstract]. Arthritis Rheum 2000;43:S319.
20. Cole A, Ong VH, Denton CP. Renal disease and systemic sclerosis: an update on scleroderma renal crisis. Clin Rev Allergy Immunol 2023; 64(3):378–391.
21. Steen VD, Medsger TA Jr. Long-term outcomes of scleroderma renal crisis. Ann Intern Med 2000;133(8):600–603.
22. Sobanski V, Dauchet L, Lefèvre G, et al. Prevalence of anti-RNA polymerase III antibodies in systemic sclerosis: new data from a French cohort and a systematic review and meta-analysis. Arthritis Rheumatol 2014;66(2):407–417.
23. Gordon SM, Stitt RS, Nee R, et al. Risk factors for future scleroderma renal crisis at systemic sclerosis diagnosis. J Rheumatol 2019;46(1): 85–92.
24. Bukamur H, Aqtash O, Shahoub I, et al. Atypical rapid onset Scleroderma Renal Crisis (SRC) complicated with diffuse alveolar hemorrhage and pleuro-pericardial effusions in a patient with recently diagnosed breast cancer and a positive anti-RNA polymerase III Ab.: A case report. Respir Med Case Rep 2018;25:137–141.
25. Shah AA, Rosen A, Hummers L, et al. Close temporal relationship between onset of cancer and scleroderma in patients with RNA polymerase I/III antibodies. Arthritis Rheum 2010;62(9):2787–2795.
26. Uddin M, Mir T, Surapaneni S, et al. Scleroderma hypertensive renal crisis among systemic sclerosis patients: a national emergency department database study. Am J Emerg Med 2022;53:228–235.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 215–226
DOI 10.1002/acr.25617

American College of Rheumatology
Empowering Rheumatology Professionals

Global and Regional Temporal Changes in Cross-Country Inequalities of Site-Specific Osteoarthritis Burden, 1990 to 2021

Haowei Chen,[1] Lei Si,[2] David J. Hunter,[3] Lei Zhang,[4] Zhe-Sheng Chen,[5] Xiaomeng Wang,[1] Vicky Duong,[6] Zhounan Zhou,[1] Peihua Cao,[1] Han Cen,[1] Su'an Tang,[1] Tianyu Chen,[7] Houjun Xue,[8] Zhaohua Zhu,[3] and Changhai Ding[9]

Objective. This study examined the global and regional temporal changes in cross-country inequalities of site-specific osteoarthritis (OA) burden from 1990 to 2021.

Methods. Age-standardized years lived with disability rates (ASYRs) for site-specific OA across 204 countries and territories were obtained from the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2021. The slope index of inequality (SII) and the concentration index (CCI) were calculated to quantify the absolute and relative cross-country inequalities. Average annual percent change (AAPC) was used to assess the temporal changes.

Results. From 1990 to 2021, the ASYR of OA increased regardless of the joint affected globally. The SII of total OA exhibited an improving cross-country inequality among higher sociodemographic index (SDI) countries and territories, decreasing from 77.26 (95% confidence interval [CI] 69.32–85.20) to 69.85 (95% CI 60.67–79.02). Similarly, the CCI of total OA decreased from 0.0815 (95% CI 0.0732–0.0897) to 0.0622 (95% CI 0.0551–0.0693). However, most regions exhibited worsening cross-country inequalities among higher SDI countries and territories. Central sub-Saharan Africa showed the largest worsening inequality in SII (AAPC 4.55, 95% CI 2.83–6.29), whereas east Asia showed the largest worsening inequality in CCI (AAPC 1.96, 95% CI 1.48–2.43). Hand OA showed consistently improving absolute and relative cross-country inequalities among higher SDI countries and territories, whereas other OA showed consistently worsening inequalities.

Conclusion. The cross-country inequalities of OA burden have persisted and even worsened in some regions over the past decades. Targeted prevention and management strategies according to geographic location and affected joint are pivotal to reduce the growing OA burden and achieve equity in health outcomes.

INTRODUCTION

Osteoarthritis (OA) is the most prevalent form of arthritis and a major cause of adult chronic pain and long-term disability.[1,2] It affected 595 million people worldwide in 2020, constituting 7.6% of the global population.[3] Notably, the prevalence of OA was higher in women than men and increased substantially with age.[3] As a major condition of substantial health care spending, OA accounted for the eighth highest amount of health care spending of $80.0 billion in the US in 2016.[4] Due to the increasing life expectancy and aging populations, the burden of OA is expected to impose a greater personal and societal toll in the foreseeable future.[5–7] However, OA is generally neglected, and the development of its treatment has not made comparable progress with other chronic diseases.[8] Up to now, there is no nonsurgical intervention that can prevent, halt, or even delay OA progression.

Supported by the National Natural Science Foundation of China (grants 82373653 and 82372428).

[1]Haowei Chen, PhD, Xiaomeng Wang, PhD, Zhounan Zhou, MBBS, Peihua Cao, PhD, Han Cen, PhD, Su'an Tang, PhD: Zhujiang Hospital of Southern Medical University, Guangzhou, China; [2]Lei Si, PhD: Western Sydney University, Sydney, New South Wales, Australia; [3]David J. Hunter, MD, PhD, Zhaohua Zhu, PhD: Zhujiang Hospital of Southern Medical University, Guangzhou, China, and Royal North Shore Hospital and The University of Sydney, Sydney, New South Wales, Australia; [4]Lei Zhang, PhD: Xi'an Jiaotong University Health Science Centre, Xi'an, China, and Melbourne Sexual Health Centre and Monash University, Melbourne, Victoria, Australia; [5]Zhe-Sheng Chen, MD, PhD: St. John's University, New York, New York; [6]Vicky Duong, PhD: The University of Sydney, Sydney, New South Wales, Australia; [7]Tianyu Chen, PhD: The Third Affiliated Hospital of Southern Medical University, Guangzhou, China; [8]Houjun Xue, MD: Foshan Sanshui District People's Hospital, Foshan, China; [9]Changhai Ding, MD, PhD: Zhujiang Hospital of Southern Medical University and Guangzhou First People's Hospital, Guangzhou, China, and University of Tasmania, Hobart, Tasmania, Australia.

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.25617).

Address correspondence via email to Changhai Ding, MD, PhD, at changhai.ding@utas.edu.au; Zhaohua Zhu, PhD, at zhaohua.zhu@utas.edu.au.

Submitted for publication January 6, 2025; accepted in revised form July 11, 2025.

SIGNIFICANCE & INNOVATIONS

- This study firstly quantified the global and regional temporal changes in absolute and relative cross-country inequalities of site-specific osteoarthritis (OA) burden over a 32-year period from 1990 to 2021.
- Despite the decline in the global cross-country inequality of OA burden, increasing cross-country inequalities were revealed in most regions from 1990 to 2021, with higher burden concentrated among countries with higher sociodemographic index (SDI) rankings.
- Hand OA consistently showed improving absolute and relative cross-country inequalities among higher SDI countries, whereas hip OA and other OA exhibited consistently worsening absolute inequalities.
- These findings underscore that progress in OA prevention and management has been uneven, and tailored strategies specific to geographic location and the affected joints are needed to effectively reduce the growing global OA burden.

The Food and Drug Administration labeling OA as a "serious disease with an unmet medical need" in 2018[9] and the World Health Organization (WHO) designating 2021 through 2030 to be the decade of healthy aging,[10] emphasizing life expectancy and quality of life, all provide opportunities to focus on OA in the context of global health.

The extent to which OA affects the population's health varies across regions and the joints involved. Age-standardized prevalence of OA in 2020 was highest in high-income North America (8.63%) and high-income Asia Pacific (8.43%) and lowest in southeast Asia (5.68%) and eastern sub-Saharan Africa (5.82%).[3] Differences in OA burden across regions might be caused by genetic, metabolic, and behavioral factors and different screening capabilities.[11–13] The knee is the most frequently affected joint, followed by the hand and the hip.[5] OA in different joints has been shown to have a distinct effect on overall health and have different pathologic mechanisms.[1,5] Taken together, a comprehensive assessment of the burden, trends, and inequalities in the distribution of site-specific OA across regions is essential for timely reorientation and enhancement of future strategies, such as adjusting the allocation of health care resources and placing more emphasis on prevention and management strategies of OA.

Past evidence has demonstrated income-related disparities in the uptake of first-line interventions for OA,[14] and patients with OA from socially disadvantaged backgrounds may not even be reached by a comparable self-management program.[15] Although previous studies have explored the burden of OA, most of them focused on the global level or a specific country, with less emphasis on cross-country inequalities at both the global and regional levels.[3,16] Additionally, most of them are limited to examining total OA or only one joint site.[17,18] To date, there remains a lack of comprehensive analysis investigating global and regional cross-country inequalities in the burden of site-specific OA from 1990 to 2021. The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2021 provided the latest estimates of a series of epidemiologic quantities of interest to quantify human health, with years lived with disability (YLDs) generally considered the best available measure of nonfatal health loss, offering an opportunity to understand the regional differences in site-specific OA burden.[19] We therefore undertook this study using GBD 2021 to quantify the cross-country inequalities in YLDs of site-specific OA with the temporal changes at global and regional levels from 1990 to 2021.

MATERIALS AND METHODS

Overview. GBD 2021 provided a comprehensive evaluation of health loss for 371 diseases and injuries using data extracted from vital registration systems, verbal autopsies, censuses, household surveys, disease-specific registries, health service contact data, and other sources.[19] The risk of bias for each data source was assessed and corrected by standardized statistical estimation using a Bayesian meta-regression tool (DisMod-MR 2.1). Detailed methods have been described extensively elsewhere.[19] This study followed the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER; Supplementary GATHER Statement).

Data source and disease definitions. GBD defined the OA reference case as symptomatic OA radiologically confirmed as Kellgren-Lawrence grade 2–4. The International Statistical Classification of Diseases and Related Health Problems (ICD) codes mapped to site-specific OA (knee, hand, hip, and other joints except for the spine) are listed in the Supplementary Appendix Methods. The age-standardized YLDs rates (ASYRs) for site-specific OA by sex across 21 GBD regions and 204 countries and territories were extracted. Because no death was attributed to OA in GBD, we used the measure of nonfatal burden, YLDs, to capture the impact on functional health and quality of life. YLDs were calculated by multiplying prevalence by the corresponding disability weight, reflecting the severity of the disability associated with the health state.[19] Then, comorbidity-adjusted YLDs were generated by simulating the distribution of all conditions and health states in the GBD cause hierarchy at the severity level and for each location-year, assuming independence. Country-level ASYR was adjusted for the country's population age structure by weighting disease burden in each age category to fit a standard population, thereby allowing for comparisons across populations with different age structures.[19] Final estimates represented the mean estimate across 500 draws, and 95% uncertainty intervals represented the 2.5th and 97.5th

percentile values across the draws. Sociodemographic Index (SDI), a composite indicator of a country's lag-distributed income per capita, average years of schooling, and the total fertility rate in women younger than 25 years, ranged from 0 (a theoretical minimum level of development relevant to health) to 100 (a theoretical maximum level of development relevant to health).[19] Countries and territories were also grouped into quintiles of high (81.03–100), high-middle (71.20–81.03), middle (61.88–71.20), low-middle (46.58–61.88), and low SDI (0–46.58), based on the 2021 values.

Statistical analysis. This study conducted a three-stage analysis. First, we described the burden of site-specific OA from 1990 to 2021 through the distribution of ASYR per 100,000 population. Pearson's or Spearman's correlation coefficients were applied to examine the possible linear or nonlinear relations between SDI and ASYR in 21 GBD regions.

Second, we calculated slope index of inequality (SII) and health inequality concentration index (CCI), two standard metrics of absolute and relative gradient inequality recommended by the WHO,[20] to quantify the cross-country inequalities in site-specific OA ASYR globally and regionally. The SII, representing the absolute difference in the burden between those with the highest and the lowest level of SDI, was calculated by regressing national OA ASYR on an SDI-associated relative position scale, which was defined by the midpoint of the cumulative class interval of the population ranked by SDI. The CCI, representing the relative difference in the burden concentrated among lower SDI or higher SDI, was calculated by numerical integration of twice the area between the concentration curve and the line of equality (45-degree line).[21] It was fitted using the cumulative fraction of OA ASYR and cumulative relative distribution of population ranked by SDI. A negative SII or CCI represents that a higher SDI corresponds to a lower ASYR, and vice versa. Larger absolute values of the SII or CCI indicate greater inequalities.

Finally, to assess the magnitude and direction of temporal trends in ASYR and inequality for OA from 1990 to 2021, we calculated average annual percent changes (AAPCs) with corresponding 95% confidence intervals (CIs) by joinpoint regression.[22] The joinpoint regression fitted the simplest joinpoint model to the data by connecting multiple distinct line segments on a logarithmic scale. Monte Carlo permutation was used to test the significance. The AAPC was calculated from the annual percent change of each segment in the last significant model in a weighted manner. If AAPC estimates and 95% CIs were both >0 (or both <0), we considered the corresponding rate to be in an upward (or downward) trend. Referring to the study by Luo et al,[23] the changing patterns of cross-country inequality were divided into six categories: (1) worsening cross-country inequality among lower SDI countries and territories, (2) improving cross-country inequality among lower SDI countries and territories, (3) worsening cross-country inequality among higher SDI countries and territories, (4) improving cross-country inequality among higher SDI countries and territories, (5) shift to higher burden among higher SDI countries and territories, and (6) shift to higher burden among lower SDI countries and territories. These categories provided a nuanced framework beyond simple increases or decreases in index values, to capture whether the nature of the inequality is evolving.[23] Their implications can be found in Supplementary Table S1.

Joinpoint analyses were conducted using Joinpoint Statistical Software version 5.3.0 (National Cancer Institute). Other statistical analyses were conducted using R software version 4.3.2 (R Project for Statistical Computing) and Stata version 17.0 (StataCorp LLC). Detailed methodologies were described in the Supplementary Appendix Methods.

Data availability statement. All data used in this research are already publicly available, which can be found from http://ghdx.healthdata.org/gbd-results-tool.

Ethics statement. This study is a secondary analysis of a publicly available database and did not involve human participants or animal subjects. Ethical approval is not applicable for this study.

RESULTS

Total OA. From 1990 to 2021, the total OA ASYR increased from 222.80/100,000 to 244.50/100,000 globally (Supplementary Table S2). In 2021, high-income Asia Pacific showed the highest ASYR (314.98/100,000), whereas southeast Asia had the lowest ASYR (196.18/100,000). The Republic of Korea, Singapore, and Brunei Darussalam showed the highest ASYR in 2021 (Supplementary Figure S1). The ASYR of total OA generally increased as the SDI value increased (Supplementary Figure S2).

Globally, significant absolute and relative inequalities in total OA ASYR were observed, with a disproportionately higher burden shouldered by countries and territories with higher SDI. From 1990 to 2021, the SII of total OA ASYR exhibited an improving inequality among higher SDI countries and territories, decreasing from 77.26 to 69.85, with an AAPC of −0.32 (95% CI −0.37 to −0.27). Similarly, the CCI decreased from 0.0815 to 0.0622, with an AAPC of −0.87 (95% CI −1.20 to −0.54; Table 1). The SII and CCI improved faster in men than women (AAPC −0.25 vs −0.19, −1.08 vs −0.39, respectively). However, most regions showed a worsening inequality among higher SDI countries and territories, with central sub-Saharan Africa showing the largest worsening inequality in SII (AAPC 4.55, 95% CI 2.83–6.29) and east Asia showing the largest worsening inequality in CCI (AAPC 1.96, 95% CI 1.48–2.43) (Supplementary Table S3, Supplementary Figure S3). Meanwhile, high-income North America presented a worsening inequality among lower SDI countries and territories for SII (AAPC −0.18, 95% CI −0.25 to −0.10). In 2021, the highest

Table 1. SII and CCI in global ASYR of site-specific OA in 1990 and 2021, with AAPC from 1990 to 2021*

	SII (95% CI)			CCI (95% CI)		
Global	1990	2021	AAPC, %	1990	2021	AAPC, %
Total OA						
Both	77.26 (69.32 to 85.20)	69.85 (60.67 to 79.02)	−0.32 (−0.37 to −0.27)	0.0815 (0.0732 to 0.0897)	0.0622 (0.0551 to 0.0693)	−0.87 (−1.20 to −0.54)
Female	94.61 (85.71 to 103.51)	89.21 (77.89 to 100.52)	−0.19 (−0.22 to −0.17)	0.0718 (0.0629 to 0.0807)	0.0637 (0.0559 to 0.0715)	−0.39 (−0.69 to −0.10)
Male	52.33 (44.12 to 60.54)	48.42 (40.40 to 56.44)	−0.25 (−0.39 to −0.10)	0.0824 (0.0730 to 0.0918)	0.0586 (0.0500 to 0.0671)	−1.08 (−1.42 to −0.75)
Knee OA						
Both	18.33 (11.26 to 25.4)	16.81 (9.01 to 24.62)	−0.29 (−0.35 to −0.24)	0.0447 (0.0304 to 0.0589)	0.0566 (0.0452 to 0.0680)	0.63 (−0.05 to 1.31)
Female	27.81 (18.31 to 37.31)	27.94 (16.83 to 39.06)	0.04 (−0.11 to 0.18)	0.0452 (0.0270 to 0.0634)	0.0733 (0.0586 to 0.0880)	1.51 (0.91 to 2.11)
Male	5.69 (0.65 to 10.72)	6.25 (0.14 to 12.37)	0.18 (−0.24 to 0.61)	0.0325 (0.0224, 0.0425)	0.0310 (0.0219 to 0.0401)	−0.20 (−1.69 to 1.32)
Hand OA						
Both	46.23 (40.99 to 51.46)	37.68 (31.71 to 43.65)	−0.66 (−0.75 to −0.58)	0.1663 (0.1382 to 0.1944)	0.0731 (0.0523 to 0.0939)	−2.67 (−3.21 to −2.12)
Female	52.16 (45.07 to 59.25)	49.09 (38.48 to 59.69)	−0.17 (−0.31 to −0.03)	0.1227 (0.0925 to 0.1528)	0.0462 (0.0228 to 0.0696)	−3.37 (−4.56 to −2.17)
Male	36.45 (32.40 to 40.51)	30.07 (25.45 to 34.69)	−0.62 (−0.72 to −0.52)	0.2171 (0.1911 to 0.2430)	0.1232 (0.1027 to 0.1436)	−1.77 (−2.06 to −1.48)
Hip OA						
Both	6.97 (5.64 to 8.3)	8.15 (6.67 to 9.63)	0.49 (0.36 to 0.61)	0.1982 (0.1681 to 0.2284)	0.1406 (0.1100 to 0.1712)	−1.15 (−1.31 to −0.98)
Female	6.07 (4.82 to 7.33)	7.27 (5.90 to 8.64)	0.61 (0.43 to 0.78)	0.1855 (0.1513 to 0.2197)	0.1319 (0.0981 to 0.1658)	−1.17 (−1.34 to −1.00)
Male	7.97 (6.60 to 9.34)	9.17 (7.61 to 10.73)	0.45 (0.36 to 0.54)	0.2111 (0.1836 to 0.2386)	0.1492 (0.1197 to 0.1787)	−1.13 (−1.35 to −0.91)
Other OA						
Both	2.27 (1.84 to 2.70)	2.79 (2.33 to 3.25)	0.62 (0.49 to 0.75)	0.0181 (0.0149 to 0.0213)	0.0202 (0.0167 to 0.0237)	0.30 (0.05 to 0.56)
Female	2.80 (2.52 to 3.07)	3.68 (3.33 to 4.03)	0.82 (0.68 to 0.96)	0.0271 (0.0241 to 0.0301)	0.0297 (0.0262 to 0.0332)	0.25 (0.00 to 0.50)
Male	2.08 (1.3 to 2.85)	2.10 (1.34 to 2.86)	−0.03 (−0.28 to 0.21)	0.0130 (0.0079 to 0.0180)	0.0119 (0.0070 to 0.0167)	−0.32 (−1.15 to 0.52)

* AAPC, average annual percent change; ASYR, age-standardized years lived with disability rate; CCI, concentration index; CI, confidence interval; OA, osteoarthritis; SII, slope index of inequality.

positive SII was observed in Oceania, and the highest positive CCI was in southern sub-Saharan Africa. The highest negative SII and the highest negative CCI were observed in high-income North America. Detailed trends of SII and CCI for total and site-specific OA from 1990 to 2021 can be seen in Supplementary Figure S4 and S5.

Knee OA. Between 1990 and 2021, the knee OA ASYR rose from 127.14/100,000 to 137.59/100,000 globally (Supplementary Table S2). High-income Asia Pacific showed the highest ASYR (180.61/100,000), and central Asia had the lowest ASYR (86.71/100,000) in 2021. The Republic of Korea, Singapore, and Brunei Darussalam emerged as the countries or territories with the highest ASYR of knee OA in 2021 (Supplementary Figure S6). Oman, Equatorial Guinea, and Thailand appeared as the countries or territories with the largest increasing ASYR of knee OA from 1990 to 2021 (Figure 1). ASYR increased as the SDI value increased (Supplementary Figure S2).

Between 1990 and 2021, the SII of knee OA ASYR showed an improving inequality among higher SDI countries and territories with the value decreased from 18.33 to 16.81, marked by an AAPC of −0.29 (95% CI −0.35 to −0.24). In contrast, the CCI of knee OA ASYR showed a worsening inequality among higher SDI countries and territories, increasing from 0.0447 to 0.0566, with an AAPC of 0.63 (95% CI −0.05 to 1.31; Table 1, Figure 2). The CCI worsened in women but remained relatively stable in men. Regionally, eastern sub-Saharan Africa reported the largest worsening inequality among higher SDI countries and territories in SII (AAPC 2.79, 95% CI 1.89–3.70). High-income North America exhibited an improving inequality among lower SDI countries and territories in the SII and CCI. In 2021, Oceania reported the highest positive SII and the Caribbean the highest positive CCI. Conversely, high-income North America showed the highest negative SII and the highest negative CCI (Figure 2, Supplementary Table S4).

Hand OA. Between 1990 and 2021, the hand OA ASYR increased from 61.67/100,000 to 70.94/100,000 globally (Supplementary Table S2). Eastern Europe reported the highest

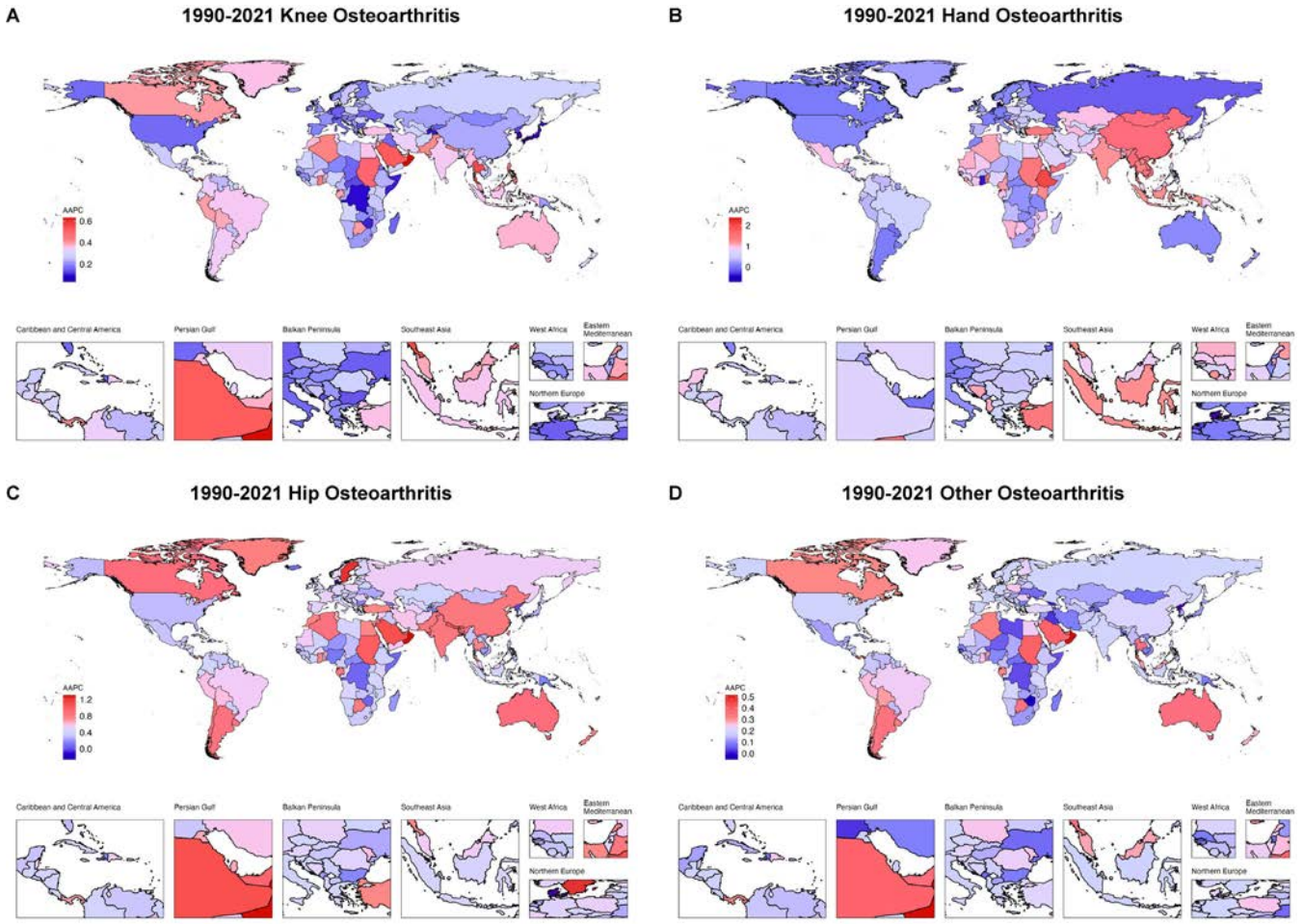


Figure 1. AAPC for ASYR of site-specific osteoarthritis from 1990 to 2021. (A) AAPC for ASYR of knee osteoarthritis from 1990 to 2021. (B) AAPC for ASYR of hand osteoarthritis from 1990 to 2021. (C) AAPC for ASYR of hip osteoarthritis from 1990 to 2021. (D) AAPC for ASYR of other osteoarthritis from 1990 to 2021. AAPC, average annual percent change; ASYR, age-standardized years lived with disability rate.

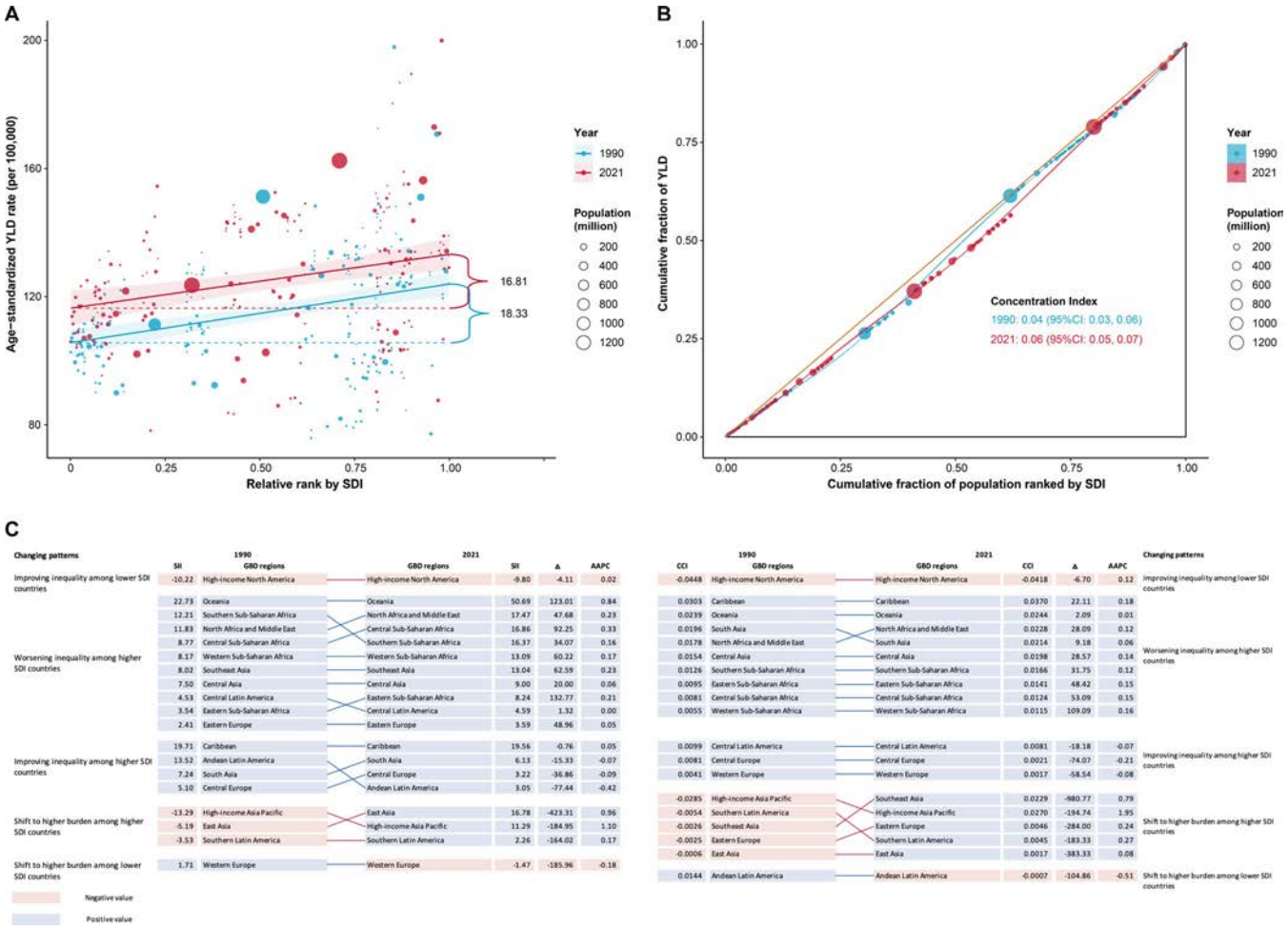


Figure 2. Absolute and relative cross-country inequalities in age-standardized YLD rate and rankings of knee OA, 1990–2021. (A) Health inequality regression curves for age-standardized YLD rate of knee OA. (B) Concentration curves for age-standardized YLD rate of knee OA. (C) Ranking graphs and change patterns for SII and CCI of knee OA. △, the percentage change of inequalities from 1990 to 2021; AAPC, average annual percent change; CCI, concentration index; CI, confidence interval; GBD, Global Burden of Diseases, Injuries, and Risk Factors Study; OA, osteoarthritis; SDI, sociodemographic index; SII, slope index of inequality; YLD, years lived with disability. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25617/abstract.

ASYR (131.95/100,000), and east Asia had the lowest ASYR (51.54/100,000) in 2021. The highest ASYRs in 2021 were in Mongolia, Russian Federation, and Estonia (Supplementary Figure S6). Equatorial Guinea, Ethiopia, and Timor-Leste emerged as the countries or territories with the largest increasing hand OA ASYR from 1990 to 2021 (Figure 1). SDI and ASYR of hand OA correlated similarly to total OA (Supplementary Figure S2).

From 1990 to 2021, the SII of hand OA ASYR demonstrated an improving inequality among higher SDI countries and territories, decreasing from 46.23 to 37.68, with an AAPC of –0.66 (95% CI –0.75 to –0.58). Similarly, the CCI decreased from 0.1663 to 0.0731, with an AAPC of –2.67 (95% CI –3.21 to –2.12) (Table 1, Figure 3). The SII showed a slower rate of improvement among women compared to men (AAPC –0.17 vs –0.62), and the CCI improved significantly faster in women than men (AAPC –3.37 vs –1.77). Regionally, Oceania presented the largest worsening inequality among higher SDI countries and territories in SII (AAPC 2.02, 95% CI –1.52 to 5.68). It is worth mentioning that high-income North America presented a worsening inequality among lower SDI countries and territories in SII (AAPC –0.13, 95% CI –0.29 to 0.03). In 2021, southern sub-Saharan Africa recorded the highest positive SII, whereas the Caribbean had the highest positive CCI. High-income Asia Pacific reported the highest negative SII and the highest negative CCI (Figure 3, Supplementary Table S5).

Hip OA. From 1990 to 2021, the ASYR of hip OA increased from 12.43/100,000 to 13.19/100,000 globally (Supplementary Table S2). High-income North America recorded the highest ASYR (28.34/100,000), whereas east Asia had the lowest ASYR (8.32/100,000) in 2021. The highest ASYRs in 2021 were seen in the US, Iceland, and United Kingdom (Supplementary Figure S6). Oman, Sweden, and Equatorial Guinea recorded the

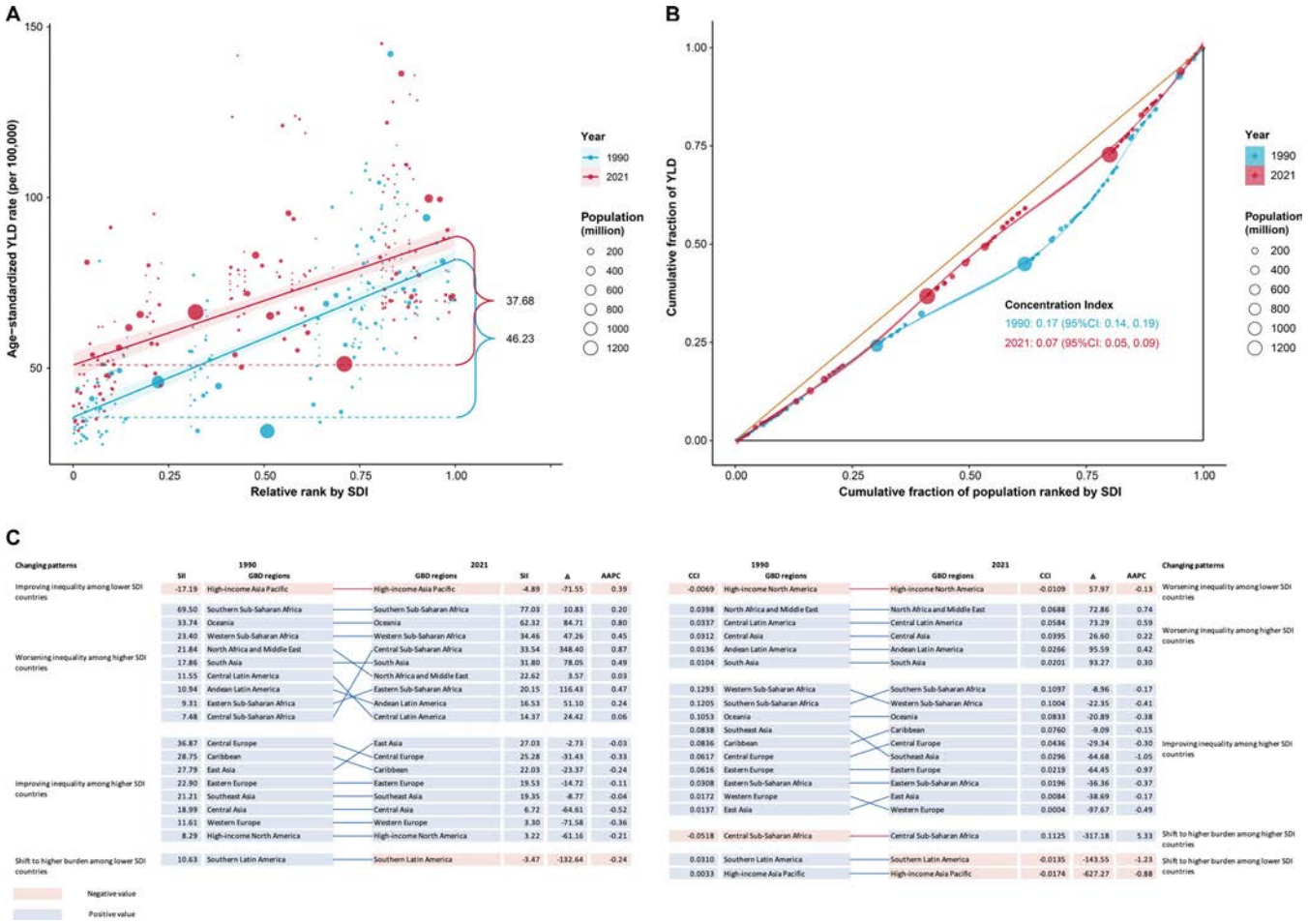


Figure 3. Absolute and relative cross-country inequalities in age-standardized YLD rate and rankings of hand OA, 1990–2021. (A) Health inequality regression curves for age-standardized YLD rate of hand OA. (B) Concentration curves for age-standardized YLD rate of hand OA. (C) Ranking graphs and change patterns for SII and CCI of hand OA. △, the percentage change of inequalities from 1990 to 2021; AAPC, average annual percent changes; CCI, concentration index; CI, confidence interval; GBD, Global Burden of Diseases, Injuries, and Risk Factors Study; OA, osteoarthritis; SDI, sociodemographic index; SII, slope index of inequality; YLD, years lived with disability. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25617/abstract.

largest increasing ASYR of hip OA from 1990 to 2021 (Figure 1). The ASYR initially slightly declined and then sharply rose as the SDI increased (Supplementary Figure S2).

Viewing from SII, the global hip OA ASYR showed a worsening inequality among higher SDI countries/territories in 1990 and 2021. The SII of hip OA ASYR increased from 6.97 to 8.15, marked by an AAPC of 0.49 (95% CI 0.36–0.61). However, the CCI of hip OA ASYR revealed an improving inequality among higher SDI countries and territories, decreasing from 0.1982 to 0.1406, with an AAPC of −1.15 (95% CI −1.31 to −0.98; Table 1, Figure 4). These trends were similar among men and women. Regionally, eastern sub-Saharan Africa exhibited the largest worsening inequality among higher SDI countries and territories in SII (AAPC 2.92, 95% CI 1.49–4.38) and the largest worsening inequality among higher SDI countries and territories in CCI (AAPC 2.89, 95% CI 1.99–3.79). High-income North America showed an improving inequality among negative SII and CCI, which still reflected a consistent burden concentration in lower SDI countries and territories. Oceania showed the highest positive SII in 2021, whereas the Caribbean had the highest positive CCI. High-income North America exhibited the highest negative SII and the highest negative CCI in the same year (Figure 4, Supplementary Table S6).

Other OA. From 1990 to 2021, the ASYR of other OA increased from 21.56/100,000 to 22.78/100,000 (Supplementary Table S2). High-income North America recorded the highest ASYR (25.20/100,000), whereas central sub-Saharan Africa recorded the lowest ASYR (20.76/100,000) in 2021. The highest ASYRs in 2021 were seen in Qatar, Kuwait, and the United Arab Emirates (Supplementary Figure S6). Oman, Equatorial Guinea, and Sudan emerged as the countries or territories with the largest increasing ASYR of other OA from 1990 to 2021 (Figure 1). Generally, the

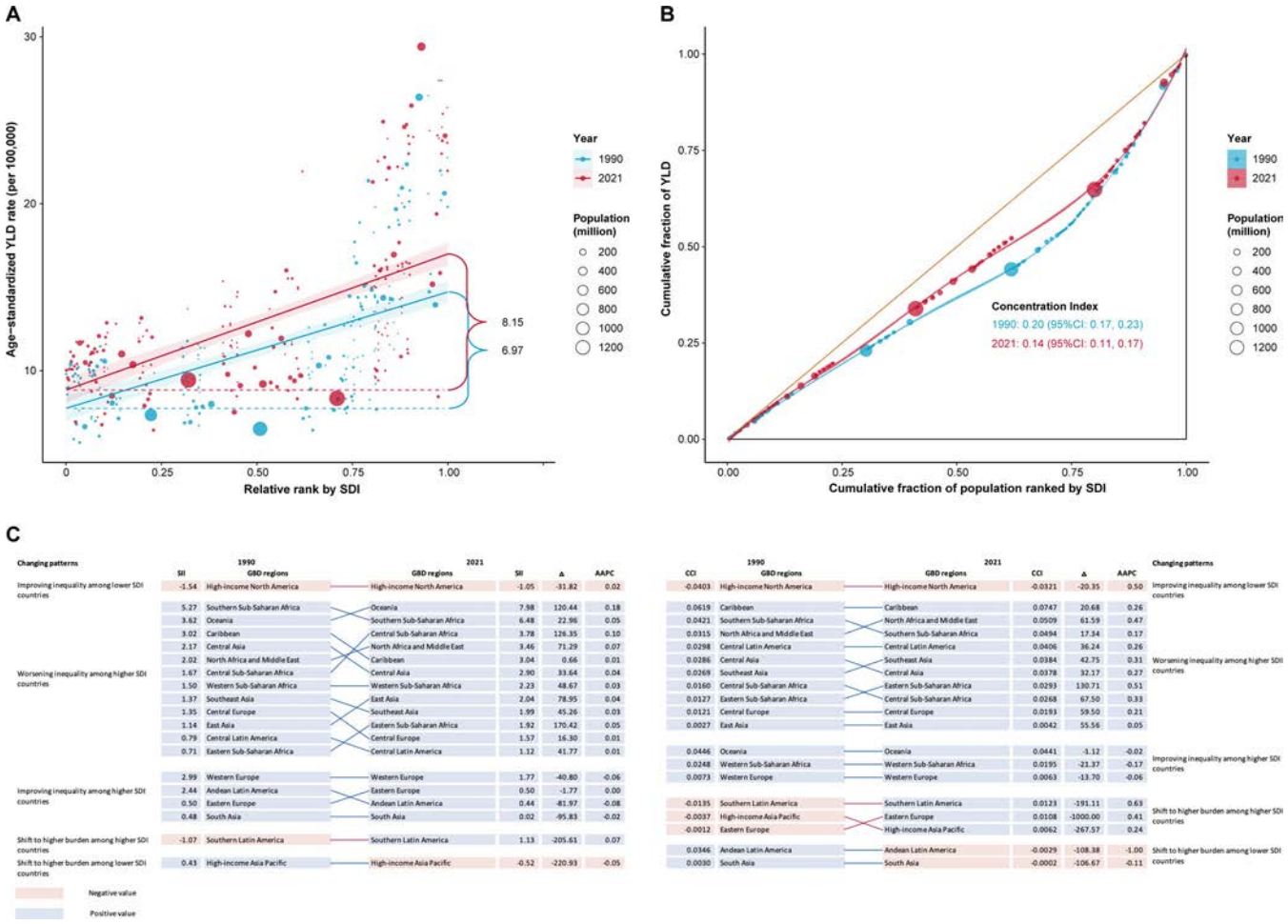


Figure 4. Absolute and relative cross-country inequalities in age-standardized YLD rate and rankings of hip OA, 1990–2021. (A) Health inequality regression curves for age-standardized YLD rate of hip OA. (B) Concentration curves for age-standardized YLD rate of hip OA. (C) Ranking graphs and change patterns for SII and CCI of hip OA. Δ, the percentage change of inequalities from 1990 to 2021; AAPC, average annual percent changes; CCI, concentration index; CI, confidence interval; GBD, Global Burden of Diseases, Injuries, and Risk Factors Study; OA, osteoarthritis; SDI, sociodemographic index; SII, slope index of inequality; YLD, years lived with disability. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25617/abstract.

ASYR initially rose and then remained stable as the SDI value increased (Supplementary Figure S2).

There showed a worsening inequality among higher SDI countries and territories in 1990 and 2021 globally. The SII of other OA ASYR increased from 2.27 to 2.79, with an AAPC of 0.62 (95% CI 0.49–0.75). Meanwhile, the CCI increased from 0.0181 to 0.0202, with an AAPC of 0.30 (95% CI 0.05–0.56; Table 1, Figure 5). Both SII and CCI worsened in women but remained relatively stable in men. Regionally, most regions showed a worsening inequality among higher SDI countries and territories in SII and CCI. Eastern sub-Saharan Africa exhibited the largest worsening inequality in SII (AAPC 3.18, 95% CI 1.95–4.42), whereas eastern Europe exhibited the largest worsening inequality in CCI (AAPC 6.90, 95% CI 5.91–7.89). High-income North America and western Europe showed an improving inequality among negative SII and CCI, which still reflected a consistent burden concentration in lower SDI countries and territories. Oceania showed the highest positive SII in 2021, whereas the Caribbean had the highest positive CCI. High-income Asia Pacific exhibited the highest negative SII and the highest negative CCI in the same year (Figure 5, Supplementary Table S7).

DISCUSSION

Using the data from GBD 2021, this study firstly quantified the global and regional cross-country inequalities in site-specific OA burden with the temporal changes from 1990 to 2021. From 1990 to 2021, an increase in global OA ASYR was demonstrated, with a disproportionately higher burden concentrated among countries and territories with higher SDI rankings. Despite the decline in the global cross-country inequality of OA burden, increasing cross-country inequalities were revealed in most

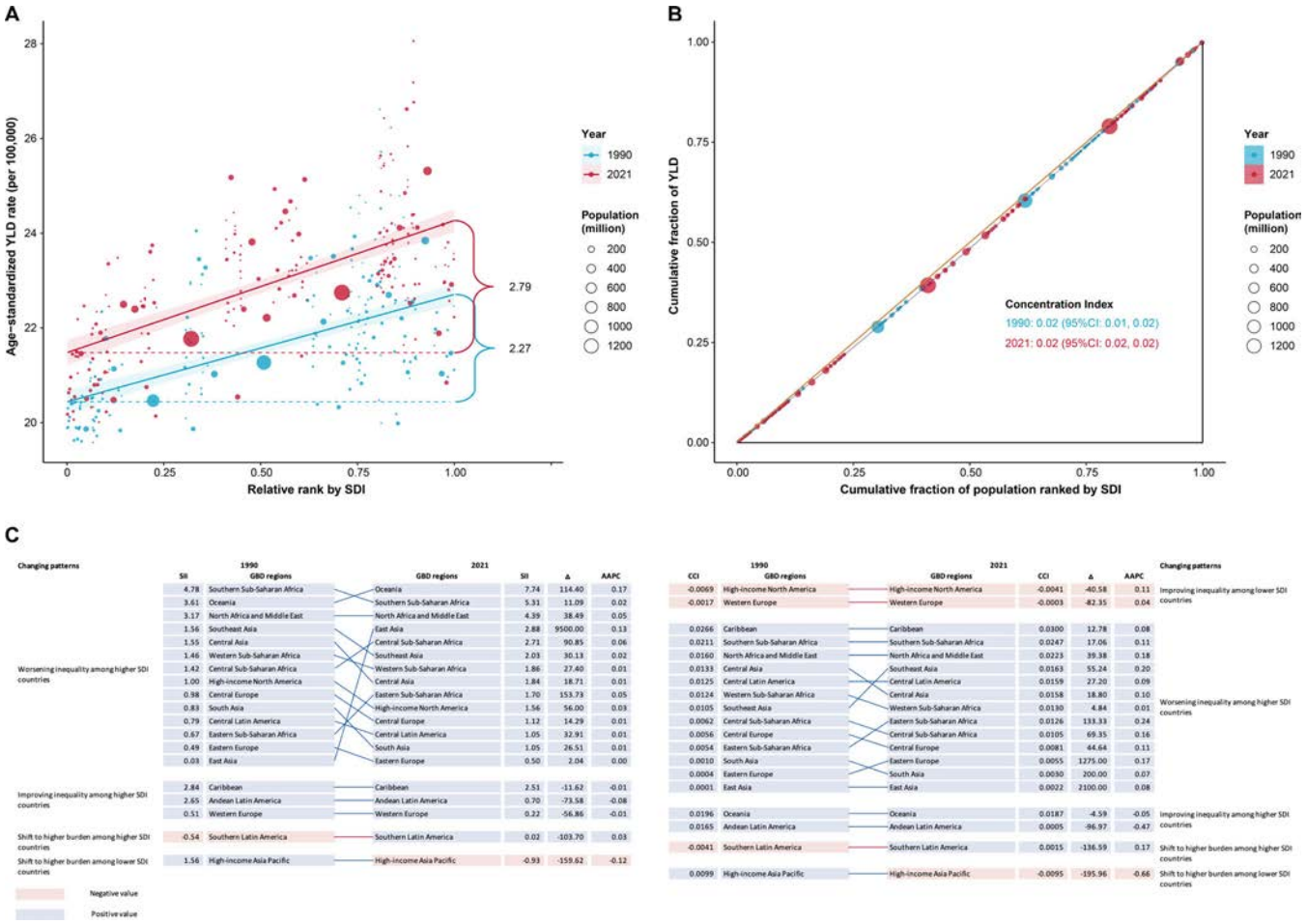


Figure 5. Absolute and relative cross-country inequalities in age-standardized YLD rate and rankings of other OA, 1990–2021. (A) Health inequality regression curves for age-standardized YLD rate of other OA. (B) Concentration curves for age-standardized YLD rate of other OA. (C) Ranking graphs and change patterns for SII and CCI of other OA. Δ, the percentage change of inequalities from 1990 to 2021; AAPC, average annual percent changes; CCI, concentration index; CI, confidence interval; GBD, Global Burden of Diseases, Injuries, and Risk Factors Study; OA, osteoarthritis; SDI, sociodemographic index; SII, slope index of inequality; YLD, years lived with disability. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25617/abstract.

regions. Central sub-Saharan Africa showed the largest worsening inequality in SII, whereas east Asia showed the largest worsening inequality in CCI. In 2021, the highest positive SII was observed in Oceania, and the highest positive CCI was in the southern sub-Saharan Africa, indicating that these regions experienced the most significant inequalities in OA burden, with substantial impact on countries and territories ranking higher in SDI. Conversely, it is worth noting that high-income North America presented a worsening inequality among lower SDI countries for SII. Among the four joint sites, hand OA had the highest SII, and hip OA had the highest CCI. There was a consistently improving inequality among higher SDI countries and territories in SII and CCI for hand OA, and other OA showed a consistently worsening inequality in SII and CCI. These findings underscore the imbalance of progress in OA prevention and management across different regions and subtypes as well as highlight the importance of addressing disparities among nations.

Globally, countries and territories with higher SDIs shouldered a disproportionately higher burden of OA. A similar study used the crude rate of disability-adjusted life-years to measure the inequality of OA at two time points (1990 and 2019) and found the inequality exacerbated over time.[16] Here, we used a stricter metrics, ASYR, to calculate the inequality of OA burden for the 32-year period from 1990 to 2021 and used AAPC to examine the temporal changes. Although inequality analysis revealed an improving inequality among higher SDI countries and territories in both the SII and CCI globally, the result of inequality was contrary to the broader understanding that higher levels of socioeconomic development correlated with better health care infrastructure, easier access to medical services, and higher overall health literacy.[24] Countries and territories with higher levels of socioeconomic development may have better availability of health services and greater capacity for early diagnosis, and the people may have a greater desire for early medical care, all of which would contribute to higher burden identified. This pattern may

also relate to economic growth, the extent increased longevity, increasing urbanization, and a parallel increase in obesity, in which lifestyle-related OA risk factors typically increase initially.[25,26] These also include higher consumption of unhealthy diets and more common sedentary lifestyles in rapidly urbanizing areas.[27,28] Although improvements in health care accessibility and quality began to emerge in these regions over time,[29] disparity was also found in terms of cross-country inequalities at the regional level. From another perspective, our analysis also revealed that the investment put into the prevention, management, and treatment of OA along with sociodemographic development may have been insufficient.[30] It is true that we must enhance economic development and improve the health care resource allocation in countries with lower SDI. At the same time, in higher SDI regions, it is also necessary to keep up with corresponding management and treatment strategies while developing the economy. To achieve this, a multifaceted approach involving various public health initiatives, targeted lifestyle interventions, and advancements in medical treatments is needed.

Notably, our analysis revealed differences in cross-country inequality trends between men and women. For instance, the absolute inequality in total OA burden showed a slower rate of improvement in women compared to men, whereas the relative inequality improved faster in women than men in hand OA. These findings suggest potential gender-specific disparities in how sociodemographic factors influence OA burden and access to care across nations and underscore the importance of considering gender dynamics when designing and prioritizing public-health initiatives aimed at reducing OA inequalities.

The worsening cross-country inequalities were found in many regions, despite global improvement. This apparent paradox likely arises from the aggregation of highly divergent regional trends. Improvements in some high-burden, high-SDI regions may have driven the global average decrease, masking significant deteriorations elsewhere. Central sub-Saharan Africa showing the largest worsening inequality among higher SDI countries and territories in SII, whereas western Europe showed the largest improving inequality in SII. This indicated that although the burden of OA remained concentrated in higher SDI countries and territories in both regions, the gap in central sub-Saharan Africa is widening significantly, whereas it is closing in western Europe. OA is a universal problem in sub-Saharan Africa. Acute and chronic locomotor problems associated with diverse entities enhance diagnostic complexity, which may lead to disparities in health care access.[31] Conversely, western Europe exhibited the lowest inequality possibly due to uniformly effective health care and chronic disease management.[32] Furthermore, high-income North America displayed a pronounced worsening inequality among lower SDI countries and territories in the SII, indicating that the disparity in OA burden between lower and higher SDI countries is widening. This can be attributed to varying degrees of economic instability and policy effectiveness, which may impact government expenditures on health and public health initiatives.[33,34] These findings collectively highlight the critical need for regionally tailored public-health policies. A "one-size-fits-all" global approach is insufficient. Regions experiencing widening gaps, particularly sub-Saharan Africa, require intensified, context-specific efforts focused on improving OA prevention (eg, addressing obesity, promoting physical activity), enhancing early diagnosis capacity and ensuring equitable access to effective management strategies. Conversely, regions showing improvement should continue and share successful models. International collaboration is vital to facilitate resource mobilization, knowledge transfer, and capacity building in regions struggling with rising inequalities.

Although the ASYR increased in all subtypes, knee OA and hand OA showed an improving inequality in SII among higher SDI countries and territories, which could largely be attributed to advancements in medical technology for these two subtypes, particularly improvements in novel approaches to managing this challenging and common condition.[35–37] The pathologic mechanism of OA is complex, and many studies were devoted to exploring the phenotypes and molecular types of knee OA and hand OA in order to better prevent and treat them.[38,39] Furthermore, inequality analysis revealed that the burden of hip OA and other OA persistently concentrated in higher SDI countries and territories, with a worsening inequality in SII. Nonpharmacological treatments are highly recommended for both knee and hip OA, but they appear to be less effective in hip OA than in knee OA.[40] In the advanced stages, these conservative approaches often fail, and total hip replacement (THR) may become necessary,[41] with 330,000 primary THRs done annually in the US.[40] However, the demand on health systems for care of patients with hip OA, including THRs, rose in all regions but might be out of reach and lead to further health inequity for individuals and countries unable to afford them.[3] The diagnosis of OA in other joint sites is often mixed with other joint diseases, which makes accurate diagnosis difficult.[42] Besides, most management measures for patients with other OA in primary care are similar to those with knee or hip OA—such as analgesia, weight management, and physiotherapy.[43] In view of this, OA in other joints requires more experienced physicians and more accurate and effective treatment; however, it may also led to significant relative inequalities as not all citizens have equal access to necessary care.

This study has several limitations. First, the data obtained from GBD rely on the modeled data because of the use of massive statistical modeling methods by GBD collaborators, particularly at the country level.[44] Differences in health care access (eg, availability of specialists, imaging technologies, and joint replacement surgery) and heterogeneous national surveillance systems may bias burden estimates. For instance, underdiagnosis in low-SDI regions likely attenuates observed inequalities, whereas overmedicalization in high-SDI regions may inflate recorded burden. Second, for Australasia and tropical Latin America, inequality

analysis is not feasible because GBD only covers two countries or territories in these regions.[23] Additionally, the hysteresis property of GBD data should be noted. Finally, our study is cross-national, which limits the understanding of disparities within individual countries with rural–urban differences.[45] Such limitations should be considered carefully when interpreting our results.

In conclusion, this study indicated that although the global cross-country inequality of OA burden decreased, increasing cross-country inequalities were revealed in most regions from 1990 to 2021, with higher SDI countries and territories disproportionately shouldering a higher burden of OA. The mismatch between OA burden management and the level of sociodemographic development, as well as the variation of the cross-country inequalities in OA burden across the joints involved, highlight an urgent need to adopt timely and effective prevention and management strategies according to geographic location and affected joint to reduce the growing OA burden and achieve equity in health outcomes.

ACKNOWLEDGMENTS

We thank the Institute for Health Metrics and Evaluation staff and its collaborators who prepared these publicly available data, and we appreciate the excellent work by the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2021 collaborators. We also thank Xiao Ming (Xiaoming_room@hotmail.com) for his help in our exploration of the GBD database. Open access publishing facilitated by University of Tasmania, as part of the Wiley - University of Tasmania agreement via the Council of Australian University Librarians.

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 authors, Drs Ding and Zhu confirm that all authors have provided the final approval of the version to be published, and take 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. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019;393(10182):1745–1759.
2. Duong V, Oo WM, Ding C, et al. Evaluation and treatment of knee pain: a review. JAMA 2023;330(16):1568–1580.
3. Steinmetz JD, Culbreth GT, Haile LM, et al; GBD 2021 Osteoarthritis Collaborators. 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.
4. Dieleman JL, Cao J, Chapin A, et al. US health care spending by payer and health condition, 1996-2016. JAMA 2020;323(9):863–884.
5. Long H, Liu Q, Yin H, et al. Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the Global Burden of Disease Study 2019. Arthritis Rheumatol 2022;74(7):1172–1183.
6. Berenbaum F. *Annals of the Rheumatic Diseases* collection on osteoarthritis (2018-2023): hopes and disappointments. Ann Rheum Dis 2024;83(2):133–135.
7. Scheuing WJ, Reginato AM, Deeb M, et al. The burden of osteoarthritis: is it a rising problem? Best Pract Res Clin Rheumatol 2023;37(2):101836.
8. Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and beyond: a Lancet Commission. Lancet 2020;396(10264):1711–1712.
9. Food and Drug Administration. Osteoarthritis: Structural Endpoints for the Development of Drugs. August 2018. Accessed December 8, 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/osteoarthritis-structural-endpoints-development-drugs.
10. World Health Organization. WHO's work on the UN Decade of Healthy Ageing (2021–2030) Accessed December 8, 2023. https://www.who.int/initiatives/decade-of-healthy-ageing.
11. Nevitt MC, Xu L, Zhang Y, et al. Very low prevalence of hip osteoarthritis among Chinese elderly in Beijing, China, compared with whites in the United States: the Beijing osteoarthritis study. Arthritis Rheum 2002;46(7):1773–1779.
12. Berenbaum F, Griffin TM, Liu-Bryan R. Review: metabolic regulation of inflammation in osteoarthritis. Arthritis Rheumatol 2017;69(1):9–21.
13. Yoshida S, Aoyagi K, Felson DT, et al. Comparison of the prevalence of radiographic osteoarthritis of the knee and hand between Japan and the United States. J Rheumatol 2002;29(7):1454–1458.
14. Battista S, Kiadaliri A, Jönsson T, et al. Income-related inequality changes in osteoarthritis first-line interventions: a cohort study. Arch Phys Med Rehabil 2024:105(3):452–460.
15. Gustafsson K, Kvist J, Eriksson M, et al. Socioeconomic status of patients in a Swedish national self-management program for osteoarthritis compared with the general population-a descriptive observational study. BMC Musculoskelet Disord 2020;21(1):10.
16. Cao F, Xu Z, Li XX, et al. Trends and cross-country inequalities in the global burden of osteoarthritis, 1990-2019: a population-based study. Ageing Res Rev 2024;99:102382.
17. Li E, Tan J, Xu K, et al. Global burden and socioeconomic impact of knee osteoarthritis: a comprehensive analysis. Front Med (Lausanne) 2024;11:1323091.
18. Zhao G, Zhu S, Zhang F, et al. Global Burden of osteoarthritis associated with high body mass index in 204 countries and territories, 1990-2019: findings from the Global Burden of Disease Study 2019. Endocrine 2023;79(1):60–71.
19. Ferrari AJ, Santomauro DF, Aali A, et al; GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2024;403(10440):2133–2161.
20. World Health Organization. Handbook on Health Inequality Monitoring: With a Special Focus on Low- and Middle-Income Countries. World Health Organization; 2013.
21. Yan W, Qin C, Tao L, et al. Association between inequalities in human resources for health and all cause and cause specific mortality in 172 countries and territories, 1990-2019: observational study. BMJ 2023;381:e073043.
22. 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.
23. Luo Z, Shan S, Cao J, et al. Temporal trends in cross-country inequalities of stroke and subtypes burden from 1990 to 2021: a secondary analysis of the global burden of disease study 2021. EClinicalMedicine 2024;76:102829.

24. Ordunez P, Martinez R, Soliz P, et al. Rheumatic heart disease burden, trends, and inequalities in the Americas, 1990-2017: a population-based study. Lancet Glob Health 2019;7(10):e1388–e1397.
25. Lim AY, Doherty M. What of guidelines for osteoarthritis? Int J Rheum Dis 2011;14(2):136–144.
26. Holowka NB, Wallace IJ, Mathiessen A, et al. Urbanization and knee cartilage growth among children and adolescents in western Kenya. ACR Open Rheumatol 2021;3(11):765–770.
27. Malik VS, Willett WC, Hu FB. Global obesity: trends, risk factors and policy implications. Nat Rev Endocrinol 2013;9(1):13–27.
28. Ng SW, Howard AG, Wang HJ, et al. The physical activity transition among adults in China: 1991-2011. Obes Rev 2014;15(S1):27–36.
29. Thomson K, Hillier-Brown F, Todd A, et al. The effects of public health policies on health inequalities in high-income countries: an umbrella review. BMC Public Health 2018;18(1):869.
30. Edwards NM, Varnum C, Nelissen RGHH, et al. The association between socioeconomic status and the 30- and 90-day risk of infection after total hip arthroplasty: a registry-based cohort study of 103,901 patients with osteoarthritis. Bone Joint J 2022;104-b(2):221–226.
31. McGill PE, Oyoo GO. Rheumatic disorders in Sub-saharan Africa. East Afr Med J 2002;79(4):214–216.
32. Seychell M. Towards better prevention and management of chronic diseases. Accessed December 8, 2024. https://ec.europa.eu/health/newsletter/169/focus_newsletter_en.htm.
33. Tucker JS, Perez LG, Klein DJ, et al. Homelessness and food insecurity during emerging adulthood: associations with changes in behavioral and physical health over a two-year period. J Adolesc Health 2024;74(5):1006–1011.
34. Blanchard L, Ray S, Law C, et al. Inequalities in research on food environment policies: an evidence map of global evidence from 2010-2020. Adv Nutr 2024;15(11):100306.
35. Wang Y, Jones G, Keen HI, et al. Methotrexate to treat hand osteoarthritis with synovitis (METHODS): an Australian, multisite, parallel-group, double-blind, randomised, placebo-controlled trial. Lancet 2023;402(10414):1764–1772.
36. Bliddal H, Bays H, Czernichow S, et al; STEP 9 Study Group. Once-weekly semaglutide in persons with obesity and knee osteoarthritis. N Engl J Med 2024;391(17):1573–1583.
37. Zhu H, Zhou L, Wang Q, et al. Glucagon-like peptide-1 receptor agonists as a disease-modifying therapy for knee osteoarthritis mediated by weight loss: findings from the Shanghai Osteoarthritis Cohort. Ann Rheum Dis 2023;82(9):1218–1226.
38. Marshall M, Watt FE, Vincent TL, et al. Hand osteoarthritis: clinical phenotypes, molecular mechanisms and disease management. Nat Rev Rheumatol 2018;14(11):641–656.
39. Lv Z, Yang YX, Li J, et al. Molecular classification of knee osteoarthritis. Front Cell Dev Biol 2021;9:725568.
40. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. JAMA 2021;325(6):568–578.
41. Learmonth ID, Young C, Rorabeck C. The operation of the century: total hip replacement. Lancet 2007;370(9597):1508–1519.
42. McCarron LV, Al-Uzri M, Loftus AM, et al. Assessment and management of ankle osteoarthritis in primary care. BMJ 2023;380:e070573.
43. Schmid T, Krause FG. Conservative treatment of asymmetric ankle osteoarthritis. Foot Ankle Clin 2013;18(3):437–448.
44. Muñoz Laguna J, Puhan MA, Rodríguez Artalejo F, et al. Certainty of the Global Burden of Disease 2019 modelled prevalence estimates for musculoskeletal conditions: a meta-epidemiological study. Int J Public Health 2023;68:1605763.
45. Long H, Xie D, Chen H, et al. Rural-urban differences in characteristics, postoperative outcomes, and costs for patients undergoing knee arthroplasty: a national retrospective propensity score matched cohort study. Int J Surg 2023;109(9):2696–2703.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 227–236
DOI 10.1002/acr.25611

Reduction in Renal Relapse and Preservation of Long-Term Kidney Function After Lupus Low Disease Activity in Patients With Lupus Nephritis

Chak Kwan Cheung,[1] Desmond Y. H. Yap,[1] Ka Lai Lee,[2] Philip H. Li,[1] Iris Y. K. Tang,[1] Chak Sing Lau,[1] and Shirley C. W. Chan[1]

Objective. Lupus low disease activity state (LLDAS) is a validated treatment target in systemic lupus erythematosus (SLE), but limited studies have explored the role of LLDAS in lupus nephritis (LN). This study aims to investigate the frequency and predictors of LLDAS attainment and its benefit on LN relapse and renal function preservation in patients with LN.

Methods. Patients with LN during 2010 to 2020 in Queen Mary Hospital and Pamela Youde Nethersole Eastern Hospital were included in the discovery cohort and validation cohort, respectively. Complete renal response (CRR), partial renal response (PRR), LLDAS, and Definition Of Remission In SLE (DORIS) remission were assessed at 12 months. Regression analysis was performed to identify risk factors of LN relapse. Receiver operating characteristic (ROC) curves were used to evaluate target attainment and long-term kidney function.

Results. A total of 245 patients with LN (discovery cohort n = 143 and validation cohort n = 102) were included. At 12 months, 57 of 143 (40%), 14 of 143 (10%), 70 of 143 (49%), and 15 of 143 (10%) patients achieved CRR, PRR, LLDAS, and DORIS remission, respectively. Attainment of both CRR/PRR and LLDAS at 12 months was associated with best relapse-free survival ($P < 0.001$). Multivariate analysis showed independent association of CRR/PRR and LLDAS with LN relapse risk reduction (CRR/PRR: hazard ratio [HR] 0.31, $P = 0.007$; LLDAS: HR 0.38, $P = 0.029$). LLDAS attainment predicts renal function preservation with satisfactory performance in both the discovery and validation cohorts (area under the curve of the ROC 0.71).

Conclusion. LLDAS is an attainable target in LN comparable to CRR/PRR. Attainment of both targets is associated with additional benefits on relapse risk reduction. Early LLDAS attainment is associated with renal function preservation.

INTRODUCTION

Lupus nephritis (LN) is an important manifestation affecting 50% to 60% of patients with systemic lupus erythematosus (SLE).[1,2] Patients with LN are at risk of disease relapse, irreversible renal damage, and subsequent progression to end-stage renal disease (ESRD).[3] Given the association between proteinuria and long-term renal outcomes, current treatment targets are primarily based on improvement in proteinuria and preservation of renal function as measured by serum creatinine or estimated glomerular filtration rate (eGFR).[4,5] Despite the usefulness of current treatment targets, a considerable number of patients experience LN relapse.[3] Other factors are believed to influence overall outcomes in LN, including serological activities and nonrenal disease activities.[6,7]

Lupus low disease activity state (LLDAS) is a validated treatment target associated with improved patient outcomes, including reduced risks of flare and prevention of organ damage in SLE.[8] LLDAS encompasses five important domains related to disease and immunosuppressive treatment. Prior studies have shown a lower LLDAS attainment among patients with SLE and renal involvement, but studies to investigate the role of LLDAS in patients with LN remain limited.[9–11] A more recent study on pediatric patients with LN demonstrated a beneficial role of LLDAS in

[1]Chak Kwan Cheung, MBBS, Desmond Y. H. Yap, PhD, Philip H. Li, MD, Iris Y. K. Tang, MBBS, Chak Sing Lau, MD, Shirley C. W. Chan, MBBS: Queen Mary Hospital, The University of Hong Kong, Pokfulam, Hong Kong; [2]Ka Lai Lee, MBChB: Pamela Youde Nethersole Eastern Hospital, Chai Wan, Hong Kong.

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.25611).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25611.

Address correspondence via email to Shirley C. W. Chan, MBBS, at scwchan@hku.hk.

Submitted for publication February 1, 2025; accepted in revised form July 7, 2025.

SIGNIFICANCE & INNOVATIONS

- Lupus low disease activity state (LLDAS) is an attainable target associated with reduced renal relapse in lupus nephritis (LN).
- Attainment of LLDAS in addition to complete and partial renal response confers extra benefit on LN relapse reduction.
- LLDAS attainment predicts long-term renal protection.

the reduction of risk of LN relapse and damage accruals.[12] Dedicated studies are needed to ascertain the benefit of LLDAS attainment and to compare its performance with current LN treatment targets.

Using longitudinal clinical data from an inception cohort of Chinese patients with LN from two tertiary hospitals in Hong Kong, this study aimed to investigate the frequency of LLDAS attainment and its potential benefit on LN relapse risk reduction and kidney function preservation in comparison with conventional renal response criteria. The study also evaluated clinical predictors of LLDAS attainment and LN relapse.

PATIENTS AND METHODS

Study population and data collection. *Discovery cohort.* The discovery cohort comprised 143 patients with biopsy-proven active LN (incident or prevalent cases) during the period of 2010 to 2020 in Queen Mary Hospital. Data from the discovery cohort were used to evaluate the frequency and predictors of LLDAS attainment as well as the effect of LLDAS attainment on the risk of relapse and kidney function preservation. Baseline demographics including sex and age of disease onset were collected. The baseline SLE Disease Activity Index 2000 (SLEDAI-2K) of all patients was documented.[13] All patients were observed regularly at intervals no longer than every 4 months for clinical and laboratory test monitoring, with additional visits decided by treating physicians depending on clinical needs. Renal biopsies were evaluated by pathologists according to the International Society of Nephrology/Renal Pathology Society 2003 or 2018 classifications.[14] Blood parameters, including serum albumin, creatinine, eGFR, serum complements, and anti–double-stranded DNA (dsDNA) titer, were documented at every visit. eGFR was calculated using the four-variable modification of diet in renal disease formula.[15] The urine profile included 24-hour urine protein or the urine protein-to-creatinine ratio (UPCR). Deidentified data from the current study will be made available upon reasonable request to the corresponding author.

The treatment regimen was decided by treating physicians (rheumatologists or nephrologists specialized in SLE) based on histologic subtypes, renal function deterioration, extrarenal activities, other comorbidities, and patient preferences. Treatment was made up of an induction phase, which included high-dose glucocorticoids and immunosuppressants, followed by maintenance therapy. First-line induction treatment for proliferative LN included mycophenolate mofetil (MMF) or cyclophosphamide, whereas calcineurin inhibitors or azathioprine were used in certain patients based on other considerations (eg, pregnancy wish) or in patients with nonproliferative subtypes.

Validation cohort. An independent validation cohort was made up of 102 patients with biopsy-proven active LN between 2010 and 2020 from Pamela Youde Nethersole Eastern Hospital. All patients in the validation cohort had regular follow-up visits to monitor treatment response and LN relapse.

Definitions of criteria. Renal response and LLDAS were assessed 12 months after the biopsy date. Renal response criteria included complete renal response (CRR), partial renal response (PRR), and no response (NR). CRR was defined as proteinuria ≤0.5 g/day with normal or near-normal eGFR (defined as within 10% of 90 mL/min/1.73m^2), PRR was defined as a reduction in proteinuria by ≥50% but >0.5 g/day with normal or near-normal eGFR, and NR was defined as not meeting either CRR or PRR. Reduced eGFR beyond near-normal range will be classified as NR regardless of the proteinuric response.[16]

LLDAS was defined as meeting all of the following criteria: (1) SLEDAI-2K score ≤4 with no activity in major organ systems and no hemolytic anemia or gastrointestinal activity, (2) no new lupus disease activity compared with the previous assessment, (3) a Safety of Estrogens in Lupus Erythematosus National Assessment–SLEDAI Physician Global Assessment (PGA) (scale 0–3) score ≤1, (4) a current prednisolone dose ≤7.5 mg/day, and (5) standard maintenance doses of immunosuppressants and approved biologic agents.[8] Definition Of Remission In SLE (DORIS) remission was defined as meeting all of the following criteria: (1) clinical SLEDAI-2k score of 0, (2) PGA < 0.5, (3) a current prednisolone dose ≤5 mg daily, and (4) stable doses of immunosuppressants and biologics.[17]

Study endpoints. The definition of LN relapse in the current study was adapted from similar or landmark trials.[18] LN relapse was suspected when there was worsening of proteinuria and/or urinary sediments and/or serum creatinine, combined with clinical judgment by the attending rheumatologist/nephrologist, and confirmed with renal biopsy. Worsening of proteinuria was defined as an increase of the UPCR or 24 hour urine protein to ≥1 mg/mg or 1 g/day in patients with ≤0.5 g at end of induction; or ≥2 mg/mg or 2 g/day in patients with >0.5 g at the end of induction. Urinary sediments included hematuria, pyuria or presence of hyaline, and granular or cellular casts. Worsening of serum creatinine was defined as increase of >25% from the lowest value after the end of induction therapy. All patients had attained clinically meaningful treatment response (defined as a reduction of 24-hour urinary protein or the equivalent of UPCR

by 50% if the baseline 24-hour urinary protein was <3 g or to <3 g if baseline proteinuria was >3 g) before renal relapse. All LN relapses were assessed after the initial 12 months.

Deterioration of renal function was defined as sustained impairment with doubling of baseline serum creatinine. Chronic kidney disease (CKD) stages were defined according to the Kidney Disease: Improving Global Outcomes definition based on eGFR.[18] CKD stage one was defined as eGFR ≥90 mL/min/1.73m^2, stage two was 60 to 89 mL/min/1.73m^2, stage three was 30 to 59 mL/min/1.73m^2, stage four was 15 to 29 mL/min/1.73m^2, and stage five was <15 mL/min/1.73m^2. Baseline demographics, laboratory parameters, and serologic markers along with the choice of induction and maintenance therapy were included in the association analyses of LLDAS attainment at 12 months and LN relapse.

Statistical analysis. The sample size required to achieve 85% power with a 5% margin of error was calculated to be 133 based on the assumption of a 20% LN relapse rate in previous publications. Continuous variables were expressed as median and interquartile range (IQR). Categorical variables were expressed as percentages. Logistic regression analysis was performed to identify predictors associated with LLDAS, and the results were reported as odds ratios (ORs) with 95% confidence intervals (CIs). Time to LN relapse between LLDAS attainment and renal response criteria was compared using the Kaplan-Meier curve and log-rank test. Cox regression analysis was performed to identify factors associated with LN relapse. Variables with a *P* value <0.10 in the univariable analysis were included in the multivariable analysis, and variables with a *P* value <0.05 in the multivariable regressions were considered statistically significant. Receiver operating characteristic (ROC) curves with area under the curve (AUC) were used to evaluate the performance of LLDAS at predicting renal function deterioration. Statistical analysis was performed on the statistical software IBM SPSS Statistics (version 28.0.1; IBM Co). A two-sided *P* value <0.05 was considered statistically significant. Survival curves were plotted with R software (version 4.2.2). The study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority of Hong Kong West Cluster.

RESULTS

The discovery cohort was made up of 143 patients with LN with a median disease duration of 14 years (IQR 7.0–20.0 years). Most patients were female (131 of 143; 92%) and had either class IV ± V (68 of 143; 48%) or class III ± V (38 of 143; 27%) LN (Table 1). Active serological activity was found in most patients, including 124 of 143 (87%) and 116 of 143 (81%) patients with hypocomplementemia and elevated anti-dsDNA titer, respectively. A small number of patients (8 of 143; 6%) had inactive serology, and renal biopsy was performed because of proteinuria and/or urinary sediments and/or impaired renal function. Most patients had preserved renal function (83 of 143; 58%) and subnephrotic-range proteinuria (116 of 143; 81%). Extrarenal disease (as measured by SLEDAI) occurred in 68 of 143 (48%) patients. Mucocutaneous (41 of 143, 29%), hematologic (25 of 143, 17%), and musculoskeletal systems (6 of 143, 4%) were the most common extrarenal domain with clinical disease activity at baseline. The median follow-up duration was 8.8 years (IQR 6.0–10.5 years), and 32 patients developed LN relapse.

LLDAS is an attainable treatment target comparable to CRR/PRR in patients with LN. The frequency and predictors of LLDAS attainment were analyzed in the discovery cohort. At 12 months, 57 of 143 (40%) and 14 of 143 (10%) patients in the discovery cohort achieved CRR and PRR, respectively. LLDAS and DORIS remission attainment was observed in 70 of 143 (49%) and 15 of 143 (9.1%) patients, respectively. Among 71 patients who attained CRR/PRR at 12 months, 40 of 71 (56%) also attained LLDAS. The remaining 42 of 143 (29%) patients in the discovery cohort failed to attain CRR/PRR or LLDAS at 12 months.

The fulfillment of each of the five LLDAS criteria was further assessed. At 12 months, the most frequently fulfilled criteria included tolerated standard therapy (140 of 143, 98%), no new disease activity (132 of 143, 92%), and PGA ≤1 (132 of 143, 92%). Only 106 of 143 (74%) and 88 of 143 (62%) of patients attained a prednisolone dose of ≤7.5 mg/day and SELDAI ≤ 4 without major organ involvement, respectively.

Baseline predictors for LLDAS attainment at 12 months in patients with active LN were evaluated in association analysis (Table 2; Supplementary Table S1). In the multivariable analysis, anti-Sm positivity was a significant negative predictor of LLDAS attainment (adjusted OR 0.33; 95% CI 0.13–0.86, *P* = 0.024). Other variables showed no association with LLDAS attainment, including age, sex, histologic class, proteinuria, serum creatinine, and choice of induction agents.

Early LLDAS and CRR/PRR attainments can be applied to LN relapse risk stratification. Among 143 patients in the discovery cohort, 32 patients developed LN relapse. The median time to LN relapse was 2.8 years (IQR 1.5–5.0 years). Patients who developed LN relapse had more frequent nephrotic-range proteinuria (LN relapse 31% vs no relapse 15%, *P* = 0.042) and hypoalbuminemia (LN relapse 29 g/L vs no relapse 32 g/L, *P* = 0.040) at baseline (Supplementary Table S2). The most frequent histologic subtypes at relapse were class IV ± V (20 of 32, 63%) and class III ± V (8 of 32, 25%). All patients were receiving glucocorticoids and maintenance immunosuppressive therapy at the time of LN relapse. Approximately 70% of patients were receiving hydroxychloroquine. Most patients had active serologic activity at the time of LN relapse (Supplementary Table S3).

Table 1. Comparison of clinical characteristics of the discovery and validation cohorts*

Baseline characteristics	Discovery cohort (n = 143)	Validation cohort (n = 102)	*P* value
Sex (female), n/N (%)	131/143 (92)	89/102 (87)	0.267
Age at SLE onset, median (IQR), years	28 (20–35)	30 (18–39)	0.495
Disease duration, median (IQR), years	14 (7–20)	6 (1–12)	<0.001
Follow-up duration, median (IQR), years	8.8 (6.0–10.5)	9.5 (6.4–10.0)	0.013
ISN/RPS LN classification, n/N (%)			
Class I/II	12/143 (8)	16/102 (16)	0.077
Class III (± V)	38/143 (27)	33/102 (32)	0.326
Class IV (± V)	68/143 (48)	40/102 (40)	0.195
Class V	25/143 (18)	13/102 (13)	0.313
24hUP (g) or UPCR (mg/mg), median (IQR)	1.6 (1.2–2.3)	2.1 (1.2–3.8)	0.006
>3 g/day, n/N (%)	27/143 (19)	34/102 (33)	0.010
Serum albumin, median (IQR), g/L	32 (28–35)	31 (26–35)	0.032
Serum creatinine, median (IQR), μmol/L	64 (52–87)	70 (60–93)	0.041
eGFR, median (IQR), mL/min/1.73m^2	98 (67–123)	92 (67–110)	0.295
CKD, n/N (%)			
CKD1	83/143 (58)	52/102 (51)	0.273
CKD2	35/143 (25)	28/102 (28)	0.599
CKD3	19/143 (13)	20/102 (20)	0.182
CKD4	3/143 (2)	1/102 (1)	0.321
CKD5	3/143 (2)	1/102 (1)	0.769
Disease factors, median (IQR)			
SLEDAI score	8 (8–11)	8.5 (8–11)	0.237
PGA score	3 (2–3)	3 (2–3)	0.246
Immunologic factors, n/N (%)			
Low C3/C4	124/143 (87)	88/102 (88)	0.724
Anti-dsDNA	116/ 143 (81)	93/102 (91)	0.028
Anti-Sm	25/136 (18)	12/102 (12)	0.163
Anti-Ro	67 (49)	46/102 (45)	0.524
Anti-La	15/136 (11)	8/102 (8)	0.410
Anti-RNP	45/136 (33)	23/102 (23)	0.075
Induction agent			
GC, n/N (%)	142/143 (99)	100/102 (98)	0.572
GC dose, median (IQR), mg	40 (30–50)	30 (25–45)	0.040
MMF, n/N (%)	111/143 (78)	66/102 (65)	0.026
CTX, n/N (%)	4/143 (3)	5/102 (5)	0.388
AZA, n/N (%)	14/143 (10)	17/102 (17)	0.110
CNI, n/N (%)	7/143 (5)	13/102 (13)	0.027
HCQ, n/N (%)	78/143 (55)	71/102 (70)	0.017
Biologics, n/N (%)[a]	5/143 (4)	0/102 (0)	0.078
Maintenance agent, n/N (%)			
MMF	107/143 (75)	49/102 (48)	<0.001
AZA	19/143 (13)	31/102 (30)	0.001
CNI	11/143 (8)	20/102 (20)	0.006
HCQ	85/143 (59)	70/102 (69)	0.115

* 24hUP, 24-hour urine protein; AZA, azathioprine; C3, complement 3; C4, complement 4; CKD, chronic kidney disease; CNI, calcineurin inhibitor; CTX, cyclophosphamide; dsDNA, double-stranded DNA; eGFR, estimated glomerular filtration rate; GC, glucocorticoid; HCQ, hydroxychloroquine; IQR, interquartile range; ISN/RPS, International Society of Nephrology/Renal Pathology Society; LN, lupus nephritis; MMF, mycophenolate mofetil; PGA, Physician Global Assessment; SLE, systemic lupus erythematosus; SLEDAI, SLE Disease Activity Index; Sm, Smith; UPCR, urine protein-to-creatinine ratio.
[a] Biologics included rituximab and belimumab; all biologics were given with background MMF.

Patients were categorized into four groups stratified by treatment target attainment at 12 months to compare the risk of LN relapse (Figure 1). The first group (NR) comprised 42 patients who attained neither CRR/ PRR nor LLDAS; the second group (LLDAS-only) comprised 30 patients who attained LLDAS without CRR/PRR at 12 months; the third group (CRR/PRR-only) comprised 31 patients who attained CRR/PRR without LLDAS; the fourth group comprised 40 patients (CRR/PRR and LLDAS) who attained both targets. In the NR group, 18 of 42 (43%) patients developed LN relapse ("high risk"). Compared with the NR group, LLDAS-only and CRR/PRR-only groups had reduced risk of LN relapse (LLDAS-only: hazard ratio [HR] 0.27, 95% CI 0.09–0.79, P = 0.017; CRR/PRR-only: HR 0.43, 95% CI 0.18–1.03, P = 0.058; "moderate risk"). The lowest relapse risk was observed in the CRR/PRR and LLDAS group (HR 0.15, 95% CI 0.04–0.50; P = 0.002; "low risk").

Kaplan-Meier curves demonstrating the effects of CRR/PRR or LLDAS attainment on time to LN relapse were included in

Table 2. Predictors of lupus low disease activity state attainment at 12 months in the discovery cohort*

Baseline characteristics	Univariable analysis		Multivariable analysis	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Female sex	1.38 (0.42–4.57)	0.599	–	–
Age at SLE onset, years	1.00 (0.97–1.03)	0.919	–	–
Prior history of LN	0.70 (0.36–1.35)	0.281	–	–
ISN/RPS LN classes				
Class III (± V)	Ref	Ref	–	–
Class IV (± V)	0.77 (0.40–1.97)	0.771	–	–
Class V	0.92 (0.34–2.53)	0.877	–	–
24hUP (g) or UPCR (mg/mg)	0.95 (0.78–1.17)	0.650	–	–
≥3 g	0.66 (0.28–1.55)	0.345	–	–
Serum albumin, g/L	1.05 (0.99–1.12)	0.136	–	–
Serum creatinine, μmol/L	1.00 (0.99–1.00)	0.221	–	–
eGFR, mL/min/1.73m^2	1.00 (0.99–1.01)	0.795	–	–
Immunologic factors				
Low C3	0.42 (0.16–1.12)	0.085	0.47 (0.17–1.31)	0.147
Low C4	0.88 (0.45–1.72)	0.708	–	–
Anti-dsDNA	0.60 (0.26–1.40)	0.237	–	–
Anti-Sm	0.34 (0.13–0.89)	0.027	0.33 (0.13–0.86)	0.024
Anti-Ro	0.84 (0.43–1.64)	0.603	–	–
Anti-La	0.92 (0.32–2.70)	0.878	–	–
Anti-RNP	0.60 (0.29–1.23)	0.163	–	–
Medications at induction				
Prednisolone dose	1.00 (0.98–1.03)	0.875	–	–
MMF	0.81 (0.37–1.77)	0.592	–	–
AZA	1.05 (0.35–3.16)	0.934	–	–
CNI	2.73 (0.51–14.6)	0.240	–	–
CTX	1.04 (0.14–7.62)	0.966	–	–
HCQ	1.23 (0.64–2.37)	0.541	–	–

* 24hUP, 24-hour urine protein; AZA, azathioprine; C3, complement 3; C4, complement 4; CI, confidence interval; CNI, calcineurin inhibitor; CTX, cyclophosphamide; dsDNA, double-stranded DNA; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; ISN/RPS, International Society of Nephrology/Renal Pathology Society; LN, lupus nephritis; MMF, mycophenolate mofetil; OR, odds ratio; Ref, reference; SLE, systemic lupus erythematosus; Sm, Smith; UPCR, urine protein-to-creatinine ratio.

Figure 2. Patients were censored at the time of event, lost to follow-up, and all-cause mortality. Attainment of both CRR/PRR and LLDAS was associated with the best outcome compared with CRR/PRR alone, LLDAS alone, or failure to attain either target (Figure 2).

LLDAS and CRR/PRR at 12 months reduce LN relapse risk by approximately 70%. Cox regression analysis was performed to identify independent factors of LN relapse in the discovery cohort (Table 3; Supplementary Table S3). In the univariate regression analysis, proteinuria ≥3 g at baseline (HR 2.22, 95% CI 1.02–4.58; *P* = 0.044), low serum albumin at baseline (HR 0.95, 95% CI 0.89–1.01; *P* = 0.094), anti-Sm positivity (HR 2.15, 95% CI 0.99–4.68; *P* = 0.054), CRR/PRR attainment at 12 months (HR 0.40, 95% CI 0.19–0.85; *P* = 0.018), and LLDAS attainment at 12 months (HR 0.27, 95% CI 0.12–0.63; *P* = 0.002) were identified. Patients with DORIS remission had a lower frequency of LN relapse, but this did not reach statistical significance (HR 0.30, 95% CI 0.04–2.20; *P* = 0.236). Anti-Sm positivity (HR 3.12, 95% CI 1.26–7.73; *P* = 0.014), CRR/PRR attainment at 12 months (HR 0.31, 95% CI 0.13–0.73; *P* = 0.007), and LLDAS attainment at 12 months (HR 0.38, 95% CI 0.16–0.91; *P* = 0.029) reduced LN relapse risk in multivariate analysis.

The validation cohort was made up of 102 patients with LN from Pamela Youde Nethersole Eastern Hospital. Patients in the validation cohort had significantly higher baseline proteinuria, anti-dsDNA titer, and serum creatinine levels. Follow-up duration was longer in the validation cohort. Medication use varied between the discovery and validation cohorts. MMF induction and maintenance were less frequent in the validation cohort. The use of calcineurin inhibitors was more frequent in both induction and the maintenance treatment in the validation cohort. Azathioprine maintenance was more frequent in the validation cohort. Other details of the validation cohort are summarized in Table 1.

Cox regression analysis was performed to identify factors of LN relapse in the validation cohort. In the univariate analysis, serum albumin (HR 1.05, 95% CI 0.99–1.10; *P* = 0.090), CRR/PRR attainment (HR 0.41, 95% CI 0.18–0.92; *P* = 0.030) and LLDAS attainment (HR 0.41, 95% CI 0.19–0.87; *P* = 0.020) were identified. In the multivariate analysis, CRR/PRR attainment and LLDAS attainment reduced LN relapse (CRR/PPR attainment: HR 0.43, 95% CI 0.19–0.97; *P* = 0.043; LLDAS attainment: HR 0.40, 95% CI 0.19–0.86; *P* = 0.018) (Supplementary Tables S3 and S4).

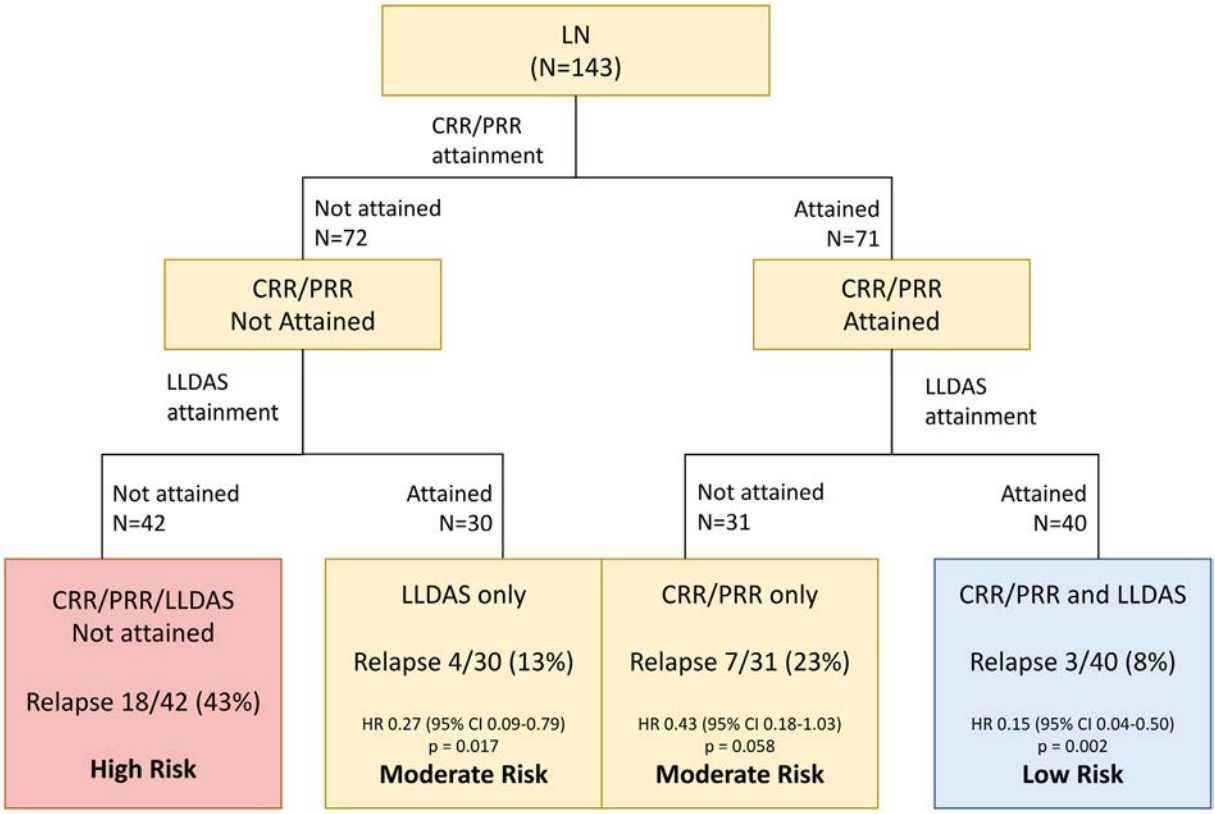


Figure 1. Relapse risk stratification based on treatment target attainment at 12 months. *Risk of renal relapse calculated using the "no response" group as the reference. CI, confidence interval; CRR, complete renal response; HR, hazard ratio; LLDAS, lupus low disease activity state; LN, lupus nephritis; PRR, partial renal response.

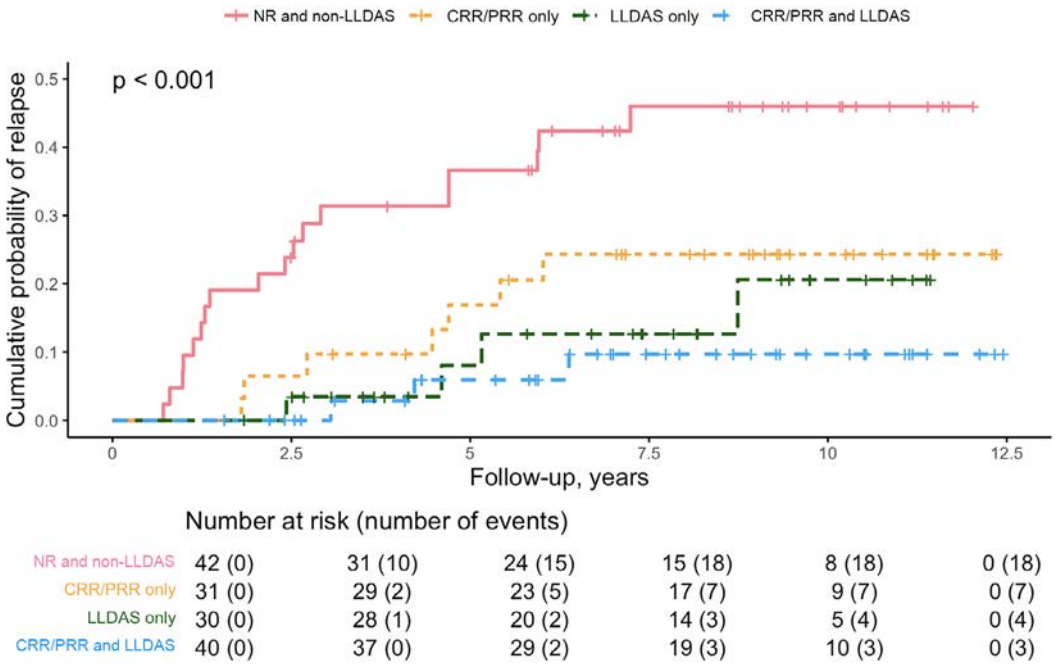


Figure 2. Cumulative risk of relapse over time stratified by renal response and LLDAS attainment. CRR, complete renal response; LLDAS, lupus low disease activity state; NR, no response; PRR, partial renal response. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25611/abstract.

Table 3. Predictors of LN relapse in the discovery cohort*

Baseline characteristics	Univariable analysis		Multivariable analysis	
	HR (95% CI)	*P* value	HR (95% CI)	*P* value
Female sex	0.56 (0.20–1.68)	0.285	–	–
Age at SLE onset, years	1.00 (0.97–1.03)	0.910	–	–
Prior history of LN	1.27 (0.63–2.58)	0.503	–	–
ISN/RPS LN classes				
Class III (± V)	Ref	Ref	–	–
Class IV (± V)	0.99 (0.46–2.18)	0.993	–	–
Class V	0.55 (0.17–1.74)	0.306	–	–
24hUP (g) or UPCR (mg/mg)	1.12 (0.95–1.31)	0.186	–	–
≥3 g	2.22 (1.02–4.58)	0.044	1.78 (0.79–4.02)	0.165
Serum albumin, g/L	0.95 (0.89–1.01)	0.094	0.97 (0.91–1.04)	0.436
Serum creatinine, μmol/L	1.00 (0.99–1.01)	0.543	–	–
eGFR (mL/min/1.73m^2)	1.00 (0.99–1.01)	0.709	–	–
Immunologic factors				
Low C3	0.95 (0.37–2.47)	0.916	–	–
Low C4	1.40 (0.67–2.91)	0.367	–	–
Anti-dsDNA	1.04 (0.43–2.55)	0.710	–	–
Anti-Sm	2.15 (0.99–4.68)	0.054	3.12 (1.26–7.73)	0.014
Anti-Ro	1.12 (0.56–2.28)	0.742	–	–
Anti-La	1.73 (0.66–4.50)	0.265	–	–
Anti-RNP	1.15 (0.55–2.40)	0.711	–	–
Induction medication				
MMF	2.15 (0.76–6.14)	0.152	–	–
AZA	0.84 (0.26–2.75)	0.772	–	–
CNI	0.05 (0–58)	0.399	–	–
CTX	1.53 (0.21–11.27)	0.676	–	–
HCQ	0.76 (0.38–1.52)	0.430	–	–
Maintenance medications				
MMF	1.18 (0.51–2.73)	0.696	–	–
AZA	1.45 (0.60–3.53)	0.411	–	–
CNI	0.04 (0–13)	0.283	–	–
Treatment targets at 12 months				
CRR/PRR	0.40 (0.19–0.85)	0.018	0.31 (0.13–0.73)	0.007
LLDAS	0.27 (0.12–0.63)	0.002	0.38 (0.16–0.91)	0.029
DORIS remission	0.30 (0.04–2.20)	0.236	–	–

* 24hUP, 24-hour urine protein; AZA, azathioprine; C3, complement 3; C4, complement 4; CI, confidence interval; CNI, calcineurin inhibitor; CRR/PRR, complete/partial renal response; CTX, cyclophosphamide; DORIS, Definition Of Remission In Systemic lupus erythematosus; dsDNA, double-stranded DNA; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; HR, hazard ratio; ISN/RPS, International Society of Nephrology/Renal Pathology Society; LLDAS, lupus low disease activity state; LN, lupus nephritis; MMF, mycophenolate mofetil; SLE, systemic lupus erythematosus; Sm, Smith; UPCR, urine protein-to-creatinine ratio.

LLDAS attainment is associated with long-term renal function preservation. Over a median follow-up duration of 8.8 years, 25 of 143 (17%) patients developed renal function deterioration (defined as sustained deterioration with doubling of baseline serum creatinine) and 13 of 143 (9%) patients developed ESRD in the discovery cohort. The AUC of LLDAS attainment for predicting renal function deterioration was 0.71 in both the discovery and validation cohorts (Figure 3).

DISCUSSION

To the best of our knowledge, this is the first study that evaluates the role of LLDAS and its usefulness in predicting the long-term LN relapse and renal function preservation in patients with LN. LLDAS is a validated treatment target in SLE associated with reduced risks of disease relapse and organ damage. Our study demonstrates that LLDAS is an attainable and beneficial treatment target in patients with LN. The frequency of LLDAS attainment is comparable to that of CRR/PRR and is associated with a reduced risk of LN relapse. The attainment of LLDAS in addition to CRR/PPR confers a greater reduction in LN relapse risk compared with the attainment of either target alone. These findings suggest that LLDAS attainment is associated with renal function preservation.

Despite treatment advances in recent decades, ESRD remains an important complication and occurs in >10% of patients with LN.[19,20] The treatment goals of LN include disease remission, prevention of disease relapse, preservation of kidney function, and minimization of drug-related toxicities.[21,22] The effectiveness of induction therapy is often assessed by surrogate parameters reflective of renal disease activity and damage, including serum creatinine, proteinuria, and urinary sediments.

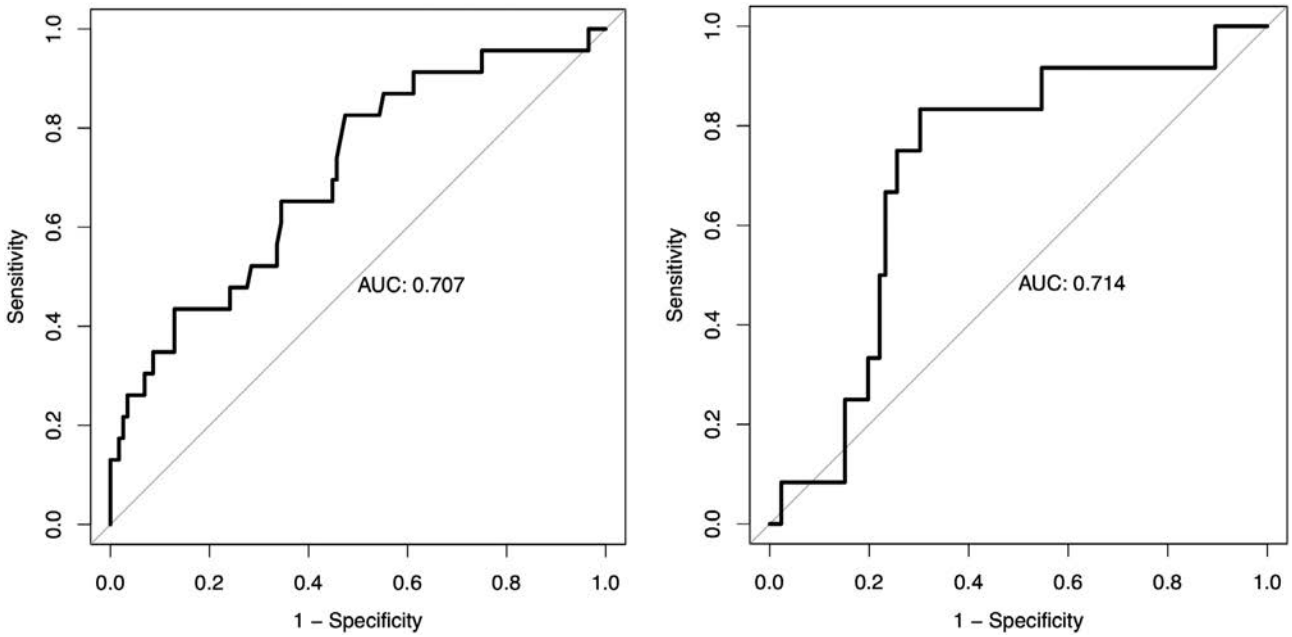


Figure 3. Receiver operating characteristic curve of LLDAS attainment for renal function deterioration in the (left) discovery cohort and (right) validation cohort. AUC, area under the curve.

CRR/PRR remains the current recommended treatment target for LN. Despite the protective benefit of CRR/PRR attainment in long-term renal outcomes, LN relapses may occur in up to half of those with LN in 10 years after the initial successful treatment.[3,23] Ongoing efforts are conducted in search of noninvasive prognostic biomarkers in LN.[24] Urinary and serum markers have shown early promise, but their applications have been mostly limited to research settings only.

Multiple studies have evaluated the usefulness of LLDAS and its association with various outcome benefits, including prevention of relapse, reduced risk of organ damage accrual, better health-related quality of life, and reduced risk of mortality. In addition to its role in defining a state of low disease activity in SLE, several studies have explored the potential application of LLDAS as a treatment target in SLE. A post hoc analysis of pooled data from the TULIP-1 and TULIP-2 trials demonstrated that LLDAS attainment was highly associated with British Isles Lupus Assessment Group-based Composite Lupus Assessment and Systemic Lupus Erythematosus Responder Index (4) responses.[25] Furthermore, anifrolumab treatment was associated with earlier, more frequent, and more prolonged and sustained LLDAS.[26] Similarly, a post hoc analysis of data from the BLISS-52 and BLISS-76 trials showed that LLDAS was useful to discriminate responders to belimumab treatment.[27] This study added evidence to the literature regarding LLDAS as a predictor of future relapse. However, the clinical application of LLDAS in LN remains to be explored. In a prospective study of 185 Chinese patients with SLE, nephritis-related markers (proteinuria and serum creatinine) and C3 levels at recruitment negatively influenced the achievement of LLDAS.[11] A similar observation was shown in a separate study of 107 White patients with SLE.[10]

SLE is a heterogeneous disease with interethnic differences in disease severity and treatment response, including more frequent and severe renal involvement among Asian patients.[28,29] One advantage of our study is the homogenous Chinese ethnicity among patients, which theoretically represents a more difficult to treat population. Furthermore, our study also showed a possible risk stratification approach based on treatment target attainment to potentially guide treatment decisions. Patients who fail to achieve CRR/PRR are known to be associated with a high risk of relapse and poor renal outcomes. Our results showed that LLDAS can be applied as an alternative treatment target with a comparable outcome on relapse prevention, and patients who attain both CRR/PRR and LLDAS represent a low-risk group with the most significant relapse risk reduction. With the ROC curve analysis illustrating a satisfactory performance of LLDAS as a predictor, our findings demonstrate promise in the potential application of LLDAS as a treatment target and endpoint for future LN trials.

Our current study failed to show an association between DORIS remission and LN relapse risk reduction largely because of the small number of patients in DORIS remission at 12 months after active LN. Post hoc analysis of major clinical trials in patients with SLE also reflected the low DORIS remission rate in patients with recent disease flare.[30] A longer time to remission is often observed compared with LLDAS, representing gradual improvement and a continuum of treatment target attainment in some patients.[31] LLDAS may therefore have a unique role in the assessment of early treatment response. A larger cohort may be

necessary to further evaluate the benefits of DORIS remission in patients with LN.

The presence of active extrarenal disease is common among patients with active LN.[32] In our study, extrarenal disease activity occurred in almost half of the patients at baseline. By incorporating the assessment of extrarenal domains (such as the SLEDAI-2K score), LLDAS has an advantage in capturing extrarenal disease activities compared with CRR/PRR. The potential benefit of LLDAS in patients with LN and extrarenal disease activity was shown in our analysis. Among patients with extrarenal disease activity at baseline, LLDAS predicted a greater reduction of relapse than CRR/PRR. This indicates that in patients with LN with extrarenal disease activity, one should not only aim for CRR/PRR in the renal domain, but also attempt to achieve LLDAS to ensure long-term disease stability. Furthermore, the degree of immunosuppressive treatments in LN warrants a delicate balance, in which inadequate immunosuppression may predispose patients to relapse while oversuppression of the immune system can lead to excessive infective complications and toxicities.[22] In this study, LLDAS and CRR/PRR showed a higher AUC than CRR/PRR alone, insinuating that LLDAS and CRR/PRR are feasible treatment targets in LN and the achievement of both targets are associated with the best outcomes on relapse prevention.

Interestingly, our study showed the association between the anti-Sm autoantibody and the risk of relapse in patients with LN. Anti-Sm is classically known to be a specific marker of SLE. Its role in patients with LN has yet to be elucidated. Previous studies showed its correlation with disease phenotypes and poorer renal outcomes.[33–35] Another study identified anti-Sm as a negative predictor for the attainment of LLDAS-50 (defined as LLDAS for ≥50% of the time).[36] Even though the current study included only a small number of patients with SLE and anti-Sm positivity, a higher risk of LN relapse and a lower likelihood of LLDAS attainment were observed among this subgroup of patients. LLDAS was shown to be helpful in this subgroup of anti-Sm–positive patients with LN in reducing the risk of relapse.

This study examined the relationship between treatment target attainment and LN relapse; future studies are needed to explore other clinical outcomes, including extrarenal flares and CKD progression. Only Chinse patients were included in our cohorts, and further studies are needed to confirm the generalizability for patients of various backgrounds. The 12-month renal response included both CRR and PRR and their effect on subsequent renal outcomes were not evaluated individually. In our cohort, most patients who attained PRR at 12 months eventually attained proteinuric target below 500 to 700 mg per day, representing a subgroup of patients with slow proteinuric response. The gradual improvement may not be applicable to all patients. Therefore, it is essential to prioritize regular monitoring of clinical progress, particularly in patients exhibiting early partial response. Prospective studies are needed to ascertain the complementary benefit of LLDAS in LN.

This study provides evidence of LLDAS as an attainable treatment target in LN. The attainment of LLDAS alone or complementary to CRR/PRR was associated with a reduced risk of LN relapse.

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 Chan 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. Chan SCW, Wang YF, Yap DYH, et al. Risk and factors associated with disease manifestations in systemic lupus erythematosus - lupus nephritis (RIFLE-LN): a ten-year risk prediction strategy derived from a cohort of 1652 patients. Front Immunol 2023;14:1200732.
2. Li PH, Wong WH, Lee TL, et al. Relationship between autoantibody clustering and clinical subsets in SLE: cluster and association analyses in Hong Kong Chinese. Rheumatology (Oxford) 2013;52(2):337–345.
3. Yap DYH, Tang C, Ma MKM, et al. Longterm data on disease flares in patients with proliferative lupus nephritis in recent years. J Rheumatol 2017;44(9):1375–1383.
4. Fanouriakis A, Kostopoulou M, Andersen J, et al. EULAR recommendations for the management of systemic lupus erythematosus: 2023 update. Ann Rheum Dis 2024;83(1):15–29.
5. Kidney Disease: Improving Global Outcomes (KDIGO) Lupus Nephritis Work Group. KDIGO 2024 Clinical Practice Guideline for the management of LUPUS NEPHRITIS. Kidney Int 2024;105(1S):S1–S69.
6. Illei GG, Takada K, Parkin D, et al. Renal flares are common in patients with severe proliferative lupus nephritis treated with pulse immunosuppressive therapy: long-term followup of a cohort of 145 patients participating in randomized controlled studies. Arthritis Rheum 2002;46(4):995–1002.
7. Kostopoulou M, Ugarte-Gil MF, Pons-Estel B, et al. The association between lupus serology and disease outcomes: a systematic literature review to inform the treat-to-target approach in systemic lupus erythematosus. Lupus 2022;31(3):307–318.
8. Franklyn K, Lau CS, Navarra SV, et al; Asia-Pacific Lupus Collaboration. Definition and initial validation of a lupus low disease activity state (LLDAS). Ann Rheum Dis 2016;75(9):1615–1621.
9. Golder V, Kandane-Rathnayake R, Hoi AY, et al; Asia-Pacific Lupus Collaboration. Frequency and predictors of the lupus low disease activity state in a multi-national and multi-ethnic cohort. Arthritis Res Ther 2016;18(1):260.
10. Piga M, Floris A, Cappellazzo G, et al. Failure to achieve lupus low disease activity state (LLDAS) six months after diagnosis is associated with early damage accrual in Caucasian patients with systemic lupus erythematosus. Arthritis Res Ther 2017;19(1):247.
11. Hao Y, Oon S, Ji L, et al. Determinants and protective associations of the lupus low disease activity state in a prospective Chinese cohort. Clin Rheumatol 2022;41(2):357–366.

12. Kisaoglu H, Baba O, Kalyoncu M. Lupus low disease activity state as a treatment target for pediatric patients with lupus nephritis. Pediatr Nephrol 2023;38(4):1167–1175.
13. Gladman DD, Ibañez D, Urowitz MB. Systemic lupus erythematosus disease activity index 2000. J Rheumatol 2002;29(2):288–291.
14. Bajema IM, Wilhelmus S, Alpers CE, et al. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. Kidney Int 2018;93(4): 789–796.
15. Levey AS, Coresh J, Greene T, et al; Chronic Kidney Disease Epidemiology Collaboration. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann Intern Med 2006;145(4):247–254.
16. Bertsias GK, Tektonidou M, Amoura Z, et al; European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association. Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of adult and paediatric lupus nephritis. Ann Rheum Dis 2012;71(11):1771–1782.
17. van Vollenhoven RF, Bertsias G, Doria A, et al. 2021 DORIS Definition Of Remission In SLE: final recommendations from an international task force. Lupus Sci Med 2021;8(1):e000538.
18. Stevens PE, Levin A; Kidney Disease: Improving Global Outcomes Chronic Kidney Disease Guideline Development Work Group Members. Evaluation and management of chronic kidney disease: synopsis of the Kidney Disease: Improving Global Outcomes 2012 clinical practice guideline. Ann Intern Med 2013;158(11):825–830.
19. Tektonidou MG, Dasgupta A, Ward MM. Risk of end-stage renal disease in patients with lupus nephritis, 1971-2015: a systematic review and Bayesian meta-analysis. Arthritis Rheumatol 2016;68(6):1432–1441.
20. Maroz N, Segal MS. Lupus nephritis and end-stage kidney disease. Am J Med Sci 2013;346(4):319–323.
21. Tamirou F, Houssiau FA. Management of lupus nephritis. J Clin Med 2021;10(4):670.
22. Lim CC, Liu PY, Tan HZ, et al. Severe infections in patients with lupus nephritis treated with immunosuppressants: a retrospective cohort study. Nephrology (Carlton) 2017;22(6):478–484.
23. El Hachmi M, Jadoul M, Lefèbvre C, et al. Relapses of lupus nephritis: incidence, risk factors, serology and impact on outcome. Lupus 2003;12(9):692–696.
24. Radin M, Miraglia P, Barinotti A, et al. Prognostic and diagnostic values of novel serum and urine biomarkers in lupus nephritis: a systematic review. Am J Nephrol 2021;52(7):559–571.
25. Morand EF, Trasieva T, Berglind A, et al. Lupus low disease activity state (LLDAS) attainment discriminates responders in a systemic lupus erythematosus trial: post-hoc analysis of the phase IIb MUSE trial of anifrolumab. Ann Rheum Dis 2018;77(5):706–713.
26. Morand EF, Abreu G, Furie RA, et al. Lupus low disease activity state attainment in the phase 3 TULIP trials of anifrolumab in active systemic lupus erythematosus. Ann Rheum Dis 2023;82(5):639–645.
27. Oon S, Huq M, Golder V, et al. Lupus low disease activity state (LLDAS) discriminates responders in the BLISS-52 and BLISS-76 phase III trials of belimumab in systemic lupus erythematosus. Ann Rheum Dis 2019;78(5):629–633.
28. Yap DY, Chan TM. Lupus nephritis in Asia: clinical features and management. Kidney Dis (Basel) 2015;1(2):100–109.
29. Petri M, Fang C, Goldman DW. East-Asian lupus nephritis in the Hopkins Lupus Cohort. Rheumatol Immunol Res 2023;4(3):157–161.
30. Parodis I, Emamikia S, Gomez A, et al. Definitions of remission in systemic lupus erythematosus: a post-hoc analysis of two randomised clinical trials. Lancet Rheumatol 2019;1(3):e163–e173.
31. Gao D, Hao Y, Mu L, et al. Frequencies and predictors of the lupus low disease activity state and remission in treatment-naïve patients with systemic lupus erythematosus. Rheumatology (Oxford) 2020; 59(11):3400–3407.
32. Carlucci PM, Preisinger K, Deonaraine KK, et al. Extrarenal symptoms associate with worse quality of life in patients enrolled in the AMP RA/-SLE Lupus Nephritis Network. Rheumatology (Oxford) 2025;64(3): 1193–1200.
33. Arroyo-Ávila M, Santiago-Casas Y, McGwin G Jr, et al. Clinical associations of anti-Smith antibodies in PROFILE: a multi-ethnic lupus cohort. Clin Rheumatol 2015;34(7):1217–1223.
34. Flechsig A, Rose T, Barkhudarova F, et al. What is the clinical significance of anti-Sm antibodies in systemic lupus erythematosus? A comparison with anti-dsDNA antibodies and C3. Clin Exp Rheumatol 2017;35(4):598–606.
35. Alba P, Bento L, Cuadrado MJ, et al. Anti-dsDNA, anti-Sm antibodies, and the lupus anticoagulant: significant factors associated with lupus nephritis. Ann Rheum Dis 2003;62(6):556–560.
36. Babaoglu H, Li J, Goldman D, et al. Predictors of predominant lupus low disease activity state (LLDAS-50). Lupus 2019;28(14):1648–1655.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 237–248
DOI 10.1002/acr.25616

American College of Rheumatology
Empowering Rheumatology Professionals

Hydroxychloroquine Associated With Lower Glomerular Filtration Rate Decline in Lupus Nephritis

Shivani Garg,[1] Brad Rovin,[2] Brad C. Astor,[1] Tripti Singh,[1] Fauzia Hollnagel,[1] Megan Kuik,[1] Lexie Kolton,[1] Callie Saric,[1] S. Sam Lim,[3] and Christie M. Bartels[1]

Objective. Hydroxychloroquine (HCQ) protects kidney function in lupus nephritis (LN) by preventing flares, yet some cohort studies show no significant benefit in kidney function with HCQ. Clarifying these conflicting findings by showing early and long-term benefits of HCQ on kidney function preservation is critical. Therefore, we analyzed data from our retrospective longitudinal inception LN cohort to examine the time-varying effects of HCQ on kidney function decline in LN.

Methods. We analyzed retrospective data from an incident biopsy-proven LN cohort. Creatinine values at LN diagnosis through the last follow-up were abstracted to find the estimated glomerular filtration rate (eGFR). Using HCQ exposure as a time-dependent covariate, we examined associations between HCQ exposure and sustained eGFR decline ≥30% and ≥40%. We also calculated an annual eGFR slope decline by HCQ exposure using linear mixed-effects analysis.

Results. Among 209 patients, 33% and 23% experienced eGFR decline ≥30% and ≥40% over time. Time-varying HCQ exposure was associated with a 60% and 62% lower risk of eGFR decline of ≥30% or ≥40%, after adjusting for propensity scores. A 77% lower risk of eGFR decline was noted in patients with chronic kidney disease (CKD) stage ≥3 with HCQ. HCQ exposure reduced the annual eGFR slope decline by 5.12 and 3.17 mL/min/1.73 m^2 within the first 5 and 10 years of diagnosis.

Conclusion. HCQ use was associated with early and long-term benefits on kidney function in LN, including those with CKD stage ≥3. Universal HCQ use should be encouraged in LN patients.

INTRODUCTION

Lupus nephritis (LN) is the most common (50%–60%) severe life-threatening manifestation of systemic lupus erythematosus (SLE).[1] Up to 30% of patients with LN develop kidney failure, and kidney failure risk is 10- to 22-fold higher (confidence interval [CI] 14–35) in patients with LN compared to age-matched peers.[2] Kidney failure often requires kidney replacement therapy, which negatively impacts the physical, social, financial, and emotional well-being of patients with LN.[2–4] Thus, preserving long-term kidney function is a major therapeutic goal in LN.[5] Hydroxychloroquine (HCQ) is a foundational therapy in SLE treatment armamentarium, with several studies noting significant improvement in disease-free and damage-free survival in SLE with HCQ. Although guidelines for LN management recommend HCQ use unless contraindicated, the strength of this recommendation was historically modest (level C).[6] More recently, the 2024 Kidney Disease Improving Global Outcomes (KDIGO) and the American College of Rheumatology (ACR) LN guidelines strongly recommend HCQ use; however, the level of evidence supporting

Presented at the American College of Rheumatology Convergence 2024, Washington, DC, November 14–19, 2024 (https://acrabstracts.org/abstract/hydroxychloroquine-users-at-lower-risk-of-kidney-function-decline-in-lupus-nephritis/).

The Office of Informatics and Information Technology was supported by the Clinical and Translational Science Award program through the National Center for Advancing Translational Sciences, NIH (grant UL1-TR-002373). This research was supported by a grant sponsored by the University of Wisconsin–Madison Institute for Clinical and Translational Research (UW ICTR). Additional support was provided by the National Center for Advancing Translational Sciences, NIH, through a Clinical and Translational Science Award to the UW ICTR (grant UL1-TR-002373). Dr Garg's work was supported by the NIH (grant 1K23-AR-084608-01). The content is solely the responsibility of authors and does not necessarily represent the official views of the funders or the NIH.

[1]Shivani Garg, MD, PhD, Brad C. Astor, PhD, MPH, Tripti Singh, MD, Fauzia Hollnagel, MPH, Megan Kuik, PharmD, Lexie Kolton, BS, Callie Saric, BS, Christie M. Bartels, MD, MS: University of Wisconsin, Madison; [2]Brad Rovin, MD: Ohio State University, Columbus; [3]S. Sam Lim, MD, MPH: Emory University, 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.25616).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25616.

Address correspondence via email to Shivani Garg, MD, PhD, at shivanigargup@gmail.com.

Submitted for publication February 25, 2025; accepted in revised form July 10, 2025.

SIGNIFICANCE & INNOVATIONS

- This study highlights kidney-specific effects of hydroxychloroquine (HCQ) exposure duration in lupus nephritis (LN) by showing a 60% and 62% lower risk of sustained estimated glomerular filtration rate (eGFR) decline of ≥30% and ≥40%, with HCQ exposure (status = yes) over time after adjusting for propensity scores.
- Particularly in patients with chronic kidney disease stage ≥3, a 77% lower risk of eGFR decline of ≥30% was noted with HCQ exposure.
- We are the first to show a yearly saving of 5.12 and 3.17 mL/min/1.73 m^2 in eGFR slope decline with HCQ within 5 and 10 years of LN diagnosis using linear mixed-effects analysis with a random slope and a random intercept.
- Finally, these findings highlight both early and long-term benefits of using HCQ in attenuating eGFR slope decline in LN and emphasize the importance of starting early and continuing HCQ therapy to preserve kidney health alongside overall lupus disease control.

this recommendation remains modest, particularly for kidney outcomes.[7,8] Moreover, Wu et al reported a higher, but not statistically significant, risk of kidney function decline in HCQ users versus nonusers (hazard ratio [HR] 1.3, CI 0.4–4.3).[9,10] Such conflicting information could potentially explain low HCQ prescribing rates (66%–79%) and even lower prescribing rates by nonrheumatologists, particularly in patients with SLE limited to the kidneys (LN).[11,12]

A few observational studies and one small randomized controlled trial (including only pediatric LN) examined the impact of HCQ on kidney outcomes (eg, LN flares, kidney failure, doubling of serum creatinine).[11,13–17] However, pooled data analysis for LN guidelines noted imprecise risk estimates and uncertainty in the current evidence.[7] This may be due to a small number of events in existing studies given these studies used terminal events (eg, kidney failure or doubling of serum creatinine), which can take a long time to occur and could be missed due to losses to follow-up or competing events. Thus, a recent consensus conference recommended using sustained kidney decline of ≥30% or ≥40% or annual estimated glomerular filtration rate (eGFR) slope decline as surrogate outcomes for eventual kidney failure to improve precision in risk estimates.[5,18,19] Moreover, data on the time-varying impact of HCQ exposure on kidney function decline have not been examined thus far.[20] Calculating an annual kidney function decline (eGFR slope decline) with and without HCQ could estimate the time-varying treatment effect of HCQ on preserving kidney function over time in LN. This information will support clinicians and patients in their decisions to commit to prolonged HCQ therapy, even those with LN and limited extrarenal manifestations.

Therefore, using a retrospective biopsy-proven inception LN cohort, we aimed to examine (1) associations between time-varying HCQ exposure and sustained eGFR decline of ≥30% and ≥40% and (2) longitudinal time-varying treatment effect of HCQ exposure on eGFR slope decline over time. Finally, we cross-checked the longitudinal time-varying treatment effect of HCQ on eGFR slope decline in a separate prevalent LN cohort identified using an electronic health records (EHR) data algorithm.

METHODS

Primary analysis using a retrospective biopsy-proven inception LN cohort. *Cohort Identification.* In this retrospective cohort study, we identified all consecutive patients with lupus who underwent native kidney biopsy between 1994 and 2019 at the University of Wisconsin Hospital and Clinics (UW) and were diagnosed as having LN. We abstracted longitudinal visit-level data since LN diagnosis until the last visit record from patients' EHR and a comprehensive kidney biopsy database. We used the standard ACR SLE classification criterion[21–23] and the 2003 International Society of Nephrology/Renal Pathology Society (ISN/RPS) LN classification to characterize patients with SLE and LN.[24] The UW Human Research Protection Program approved this study with a waiver of informed consent (institutional review board numbers 2022-0840 and 2016-1260).

Variables. Sociodemographics & comorbidities. Using the EHR, we recorded sociodemographics at the time of biopsy (LN diagnosis) including age, sex, race, insurance type (commercial vs Medicaid or no insurance), and area deprivation index (ADI) derived from nine-digit zip code (categorized as >80 vs ≤80; ADI >80 indicates neighborhoods with high socioeconomic status disadvantage).[25] Data on comorbidities including hypertension (HTN), diabetes mellitus (DM), and hyperlipidemia, at or within the year of LN diagnosis, were abstracted, consistent with published literature.[26] A standard validation method was used to ascertain comorbidities (HTN, DM, dyslipidemia), which included the following steps: (1) two or more encounters with International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, (ICD-10) codes for the comorbidity documented ≥30 days apart AND (2) two abnormal test results for the comorbidity in two or more encounters OR (3) medication prescribed for the comorbidity.

LN biopsy findings. LN classes I through VI were abstracted from the initial kidney biopsy reports.[24] LN class was categorized as proliferative, mixed (eg, class III ± V or IV ± V), and nonproliferative (eg, membranous, class I or II with or without podocytopathy). The NIH LN chronicity and activity indices were abstracted from the index LN biopsy reports.[24] Renal arteriosclerosis was graded as percent luminal narrowing.[27]

LN disease course. We abstracted visit-level data for laboratory measurements, such as serum creatinine and urine protein creatinine ratio (UPC), at LN diagnosis (at biopsy) until the last follow-up visit records available in the EHR. We found the eGFR for each visit using the National Kidney Foundation recommended race-neutral eGFR calculator (Chronic Kidney Disease Epidemiology Collaboration) and the abstracted serum creatinine values. Using KDIGO definitions for renal remission, we identified patients who achieved renal remission at 12 months after LN diagnosis.[7] Additionally, using standard definitions of LN flare,[7] we identified patients who experienced at least LN flare after diagnosis. SLE duration before LN diagnosis was categorized as less than two years or two or more years.

Time-varying HCQ exposure. We used visit-level medication prescription data from the EHR to ascertain HCQ use status (yes or no) at each time point. HCQ exposure status was ascertained as yes when there was an active HCQ prescription for at least three months at each follow-up time point. Only the prescribed dose of HCQ was assessed in this paper; data on number of days covered by prescriptions were not available. Any change in HCQ exposure status at any given time point for each patient was accounted in analysis.[28] To test the effect of a higher than guideline-recommended HCQ dose, data on median HCQ dose over time were included as >5 versus ≤5 mg/kg/day categories.

Other medications. We also abstracted data on angiotensin converting enzyme inhibitor (ACE-i) or angiotensin receptor blocker (ARB), aspirin, statin, and sodium–glucose cotransporter inhibitor ever use (yes/no). Next, we abstracted data on the type of induction and maintenance LN therapy regimen (cyclophosphamide vs mycophenolate vs combination therapies or others, such as belimumab, rituximab, calcineurin inhibitors, or azathioprine) and the duration of mycophenolate maintenance therapy (in years). Finally, glucocorticoid data were included as two categorical variables: (1) high-dose glucocorticoids (60 mg or higher) at LN diagnosis (yes/no) and (2) prolonged glucocorticoid use defined as prednisone >7.5 mg (or equivalent doses of other steroids) for more than six months (yes/no).[29]

Outcomes. We analyzed three key outcomes.

1. Primary outcome: eGFR decline of ≥30%, defined as ≥30% reduction in eGFR for at least two follow-up visits compared to baseline eGFR (at or within three months of LN diagnosis; a median value was calculated if multiple values were noted) or if a patient required sustained renal replacement therapy at any follow-up time.[5,18,19]
2. Secondary outcome: eGFR decline of ≥40%, defined as 40% or more reduction in eGFR for at least two follow-up visits from baseline or if a patient required sustained renal replacement therapy at any follow-up time.[5,18,19] The time to sustained eGFR decline of ≥30% or ≥40% was defined as time from LN diagnosis until the first follow-up visit with ≥30% or ≥40% decline in eGFR from baseline.
3. Annual eGFR slope: visit-level data on eGFR were used to estimate annual eGFR slope over follow-up. The annual eGFR slope decline usually reflects the benefits of treatment and accounts for impact of treatment (HCQ) over time. All outcomes were abstracted by independent abstractors who were blinded to other variables (CS, LK, MK).

Analysis. Descriptive data. The distribution of patient characteristics was examined. Based on distribution, continuous variables were expressed as mean ± SD, and categorical variables were expressed as n (%).

Data cleaning (missing data and censoring). Given that we abstracted visit-level data from patients' EHR, we ascertained loss of follow-up if a patient had no visit-level data, and these patients were censored from the analysis. Patients without sustained eGFR decline were censored at the end of the follow-up. Visit-level eGFR values, when used as an outcome, were not imputed. Standard methods, such as imputing median values for laboratory measurements, medications, and other variables, were used to handle missing data.[30]

Cox proportional hazards analysis using eGFR decline ≥30% as the primary outcome. We used a Cox proportional hazards (PH) model to analyze the time-varying effects of HCQ use on eGFR decline of ≥30% (primary outcome). Both unadjusted and adjusted Cox regression analyses were conducted to examine associations between time-dependent HCQ exposure and eGFR decline of ≥30%.[5,18,19,28] The fully adjusted Cox model included a time-dependent HCQ exposure variable and all relevant covariates that could influence eGFR decline. Schoenfeld residual analysis was performed, and variables such as insurance and history of HTN were stratified in the full Cox model to satisfy the PH assumption. Given the limited number of eGFR events, we also developed a reduced model that included key covariates: age, sex, race, insurance history of HTN (yes/no), baseline eGFR, NIH LN chronicity and activity indices, prolonged glucocorticoid use (yes/no), HCQ use status, and weight-based HCQ dose categories (>5 mg/kg/day vs ≤5 mg/kg/day). Race and insurance were stratified to meet PH assumptions. *E* values for the effect size and 95% CIs were calculated.[31] The *E* value quantifies the minimum strength of association that an unmeasured confounder would need to have with both the exposure and the outcome for the observed association to be null. Higher *E* values (eg, >2) indicate that a strong unmeasured confounder would be needed to negate the observed effect, thereby supporting the robustness of the findings.

Cox PH analysis using eGFR decline ≥40% (secondary outcome). Similar methods as describe in the primary outcome section were used to examine the association between time-dependent HCQ exposure and eGFR decline of ≥40%, adjusting for all covariables in the full model and only key variables in the

limited model. Again, *E* values were calculated to account for unmeasured confounders.

Inverse probability weighting–adjusted models. To control for survival bias, we developed inverse probability weighting (IPW)-adjusted models for both primary and secondary analyses. The propensity score was estimated using a logistic regression model with the following covariates: age, sex, race, insurance, history of HTN, baseline eGFR, NIH LN chronicity and activity indices, prolonged glucocorticoid use, mycophenolate mofetil (MMF) duration (in years), LN class, ACE-i/ARB ever use, HCQ use at baseline (yes or no), and median HCQ dose categories. The weights were then used in the reduced Cox PH model to separately analyze the time to eGFR decline of ≥30% or ≥40%.

Sensitivity analysis. We performed the following sensitivity analyses: (1) excluding patients with follow-up duration shorter than six months and (2) including all patients with at least 5 and 10 years of follow-up. Doing so helped us examine early effects of HCQ use and have more uniformity in the cohort's follow-up time. Additionally, variables such as LN remission and LN flares can have causal links with eGFR decline. Thus, these were not included in the primary and secondary analyses described earlier to avoid underestimation of treatment effect. However, these variables were added in separate models as part of the sensitivity analysis. Finally, we examined associations between time-dependent HCQ exposure and eGFR decline of ≥30%, including only patients with eGFR ≤60 mL/min/1.73 m^2 (chronic kidney disease [CKD] stage ≥3) at LN diagnosis.

Linear mixed-effects model to estimate annual eGFR slope decline. We used a linear mixed-effects model with random intercepts and slopes (via the `lme4` package in R [The R Project for Statistical Computing]) to analyze the time-varying effect of HCQ on eGFR slope over time. The model included fixed effects for age, sex, race, history of HTN, NIH LN chronicity and activity indices, nephrotic syndrome at LN diagnosis (defined as UPC ≥3.5), CKD stage ≥3 at LN diagnosis, ACE-i/ARB use (ever), time-varying HCQ exposure, and time. Random effects were specified for both the intercept and the slope of time, allowing for individual variability in baseline eGFR and the rate of change over time. An interaction term between time and HCQ exposure was included to assess the time-varying effect of treatment. To have more uniformity in the cohort's follow-up time and examine early effect of HCQ exposure, follow-up time for this model was limited to 10 years. To evaluate model fit and compare the inclusion of the time-varying HCQ effect, we performed a likelihood ratio test between models with and without the interaction term. A visualization of eGFR trajectories for HCQ exposure (yes vs no) was generated using the `ggplot2`, `cowplot`, and `effects` packages in R.

Additional analysis in a prevalent cohort to test treatment effect of HCQ on EGFR slope. *Prevalent LN cohort identification.* Prevalent LN cases defined as those with two or more visits at least 30 days apart with an ICD-10 code for LN (M32.14 or M32.15; 96% specificity) were identified in the EHR, and data were abstracted using an automated data query algorithm.[32]

Key variables. Using the definitions defined earlier, we abstracted data on sociodemographics (age, race, ethnicity, sex, insurance), comorbidities (history of HTN, DM, dyslipidemia), time-varying HCQ exposure using prescription dates for HCQ, ACE-i/ARB ever use, and laboratory values (creatinine to calculate eGFR and UPC to estimate nephrotic range proteinuria). Data for LN biopsy findings and LN disease course (induction regimen, maintenance regimen, and remission) were not available for prevalent cases.

Linear mixed-effects analysis to cross-check time-varying treatment effect of HCQ on annual eGFR slope. We used a linear mixed-effects model with random intercepts and slopes (as described earlier; lme4 in R) to analyze the time-varying treatment effect of HCQ on eGFR slope over time, adjusting for sociodemographics (age, race, ethnicity, sex, insurance), ACE-i/ARB, nephrotic range proteinuria, and history of HTN.[5,33]

RESULTS

Primary analysis using our retrospective inception LN cohort, N = 209. *Retrospective inception LN cohort characteristics.* Overall, 209 adults with biopsy-proven incident LN were included in the cohort with a mean duration of follow-up of 5.4 ± SD 3.8 years (Table 1). A total of 66 and 48 patients experienced sustained eGFR decline ≥30% or ≥40% over time; 35 patients experienced eGFR decline of ≥30% by year five. Key patient characteristics are highlighted in Table 1.

Time-varying HCQ exposure and eGFR decline of ≥30% (primary outcome). Using unadjusted univariable Cox regression models and HCQ exposure as a time-dependent covariate, HCQ exposure (status = yes) was associated with a 50% lower rate of sustained eGFR decline of ≥30% (unadjusted HR 0.50 [95% CI 0.29–0.86]; *P* = 0.012; Supplementary Table 1). After adjusting for covariables, time-varying HCQ exposure status (yes) was associated with 59% (adjusted HR 0.41 [95% CI 0.22–0.78]; Table 2A) and 64% (adjusted HR 0.36 [95% CI 0.19–0.68]; Table 2B) lower rates of achieving the primary outcome (eGFR decline of ≥30%) in the full and reduced models, respectively. Given sample size limitations and no association between LN specific treatments and eGFR decline in unadjusted models (Supplementary Table 1), LN specific treatments were not included in the adjusted models. Additionally, an *E* value of 5.0 was calculated for the reduced model, indicating that any unmeasured confounder needs to be associated with five-fold higher eGFR decline of ≥30% for the association with HCQ exposure to be truly null in the reduced model. Even after adjusting for propensity score, the risk of eGFR decline of ≥30% significantly

Table 1. Baseline patient characteristics (N = 209)*

Variables	Value
Age at diagnosis, mean ± SD, y	38 ± 15
Sex, n (%)	
Female	156 (75)
Male	53 (25)
Race and ethnicity, n (%)	
Asian	30 (14)
Black	24 (12)
Hispanic	15 (7)
White	140 (67)
Medicaid or no insurance, n (%)	72 (35)
ADI ≥ 80, n (%)	27 (13)
eGFR at LN diagnosis, mean ± SD, mL/min/1.73 m^2	80 ± 37
SLE duration ≥2 y before LN diagnosis, n (%)	59 (28)
HTN at/within 1 y of diagnosis, n (%)	82 (39)
Diabetes at/within 1 y of diagnosis, n (%)	11 (5)
UPC at LN diagnosis, mean ± SD, mg/mg	2.9 ± 3.3
LN class per index LN biopsy, n (%)	
Proliferative	108 (52)
Nonproliferative	55 (26)
Mixed (eg, LN class IV/V)	46 (22)
LN activity index per index LN biopsy, mean ± SD	2.5 ± 3.3
LN chronicity index per index LN biopsy, mean ± SD	2.2 ± 2.1
Renal arteriosclerosis in % luminal area narrowing, mean ± SD	22 ± 28
Renal remission at 12 mo, n (%)	
Complete remission	144 (69)
Partial remission	33 (16)
High dose of steroids at diagnosis, n (%)	112 (50)
Induction therapy after diagnosis, n (%)	
Cyclophosphamide	44 (21)
Mycophenolate	146 (70)
Combination or others	19 (9)
Maintenance therapy after diagnosis, n (%)	
Mycophenolate	155 (74)
Azathioprine or others	54 (26)
Mycophenolate use duration, mean ± SD, y	3.7 ± 3.2
Chronic steroid (≥7.5 mg/d for ≥6 mo), n (%)	88 (42)
HCQ exposure status = yes at diagnosis, n (%)	166 (79)
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes, n (%)	39 (19)
ACE-i/ARB use, n (%)	140 (67)
SGLT2i use, n (%)	13 (6)
Follow-up time, mean ± SD, y	5.4 ± 3.8
Follow-up time >5 y, n (%)	108 (44)

* eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration formular. ACE-i, angiotensin converting enzyme inhibitor; ADI, area deprivation index; ARB, angiotensin receptor blocker; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; HTN, hypertension; LN, lupus nephritis; SGLT2i, sodium–glucose cotransporter 2 inhibitor; SLE, systemic lupus erythematosus; UPC, urine protein creatinine.

decreased with HCQ exposure (IPW adjusted HR 0.40 [95% CI 0.23–0.69]; P = 0.001; Table 3A).

HCQ exposure and eGFR decline of ≥40% (secondary outcome). In an unadjusted Cox model, time-varying HCQ exposure was associated with a 58% lower likelihood of sustained eGFR decline of ≥40% (unadjusted HR 0.42 [95% CI 0.23–0.77]; P = 0.005; Supplementary Table 2). After adjusting for known covariables, HCQ exposure remained strongly associated with a 66% and 67% lower risk of eGFR decline of ≥40% in both full and reduced Cox models (Table 4). An E value of 5.5 was calculated for the reduced model. Finally, a strong association between HCQ exposure and eGFR decline of ≥40% remained even after adjusting for propensity scores (IPW-adjusted HR 0.38 [95% CI 0.21–0.70]; P = 0.002; Table 3B).

Sensitivity analysis. All patients had at least six months of follow-up. No significant changes were noted in associations between time-varying HCQ exposure with eGFR decline of ≥30% after censoring time to 5 and 10 years after LN diagnosis (Supplementary Table 3A–B).

Even after adjusting for LN remission and flares, time-varying HCQ exposure was associated with a 63% lower risk of eGFR decline of ≥30% (adjusted HR 0.37 [95% CI 0.31–0.47]; P = 0.005; not shown in a table). Finally, in a subgroup analysis

Table 2. Adjusted Cox regression model showing associations between time-varying HCQ exposure and eGFR decline of ≥30% (N = 209)*

Variables	Adjusted HR (95% CI)[a]	*P* value
Full multivariable Cox model[a]		
Age in 1-y increments	1 (0.98–1.02)	0.70
Female	0.81 (0.43–1.52)	0.51
White race	ref	
Asian race	1.88 (0.80–4.43)	0.15
Black race	**4.20 (1.86–9.48)**	**0.0005**
Hispanic ethnicity	0.76 (0.17–3.32)	0.72
eGFR per 10 mL/min/1.73 m^2 decrease	1.09 (0.98–1.20)	0.11
NIH LN chronicity index per 1-point increase	**1.28 (1.10–1.50)**	**0.002**
NIH LN activity index per 1-point increase	**1.12 (1.04–1.22)**	**0.005**
Renal arteriosclerosis per 1% increase	1.01 (0.99–1.02)	0.39
MMF exposure duration per 1-y increase	0.94 (0.86–1.03)	0.20
ACE-i/ARB ever use, yes	0.87 (0.49–1.53)	0.62
Prolonged glucocorticoids use (>7.5 mg >6 mo), yes[c]	1.74 (0.96–3.16)	0.07[d]
HCQ exposure (status = yes)[e]	**0.41 (0.22–0.78)**	**0.007**
Higher than guideline recommended HCQ dose >5 mg/kg/d, yes[f]	1.08 (0.55–2.23)	0.83
Reduced multivariable Cox model[b]		
Age in 1-y increments	1.01 (0.99–1.03)	0.40
Female	0.87 (0.47–1.62)	0.67
History of HTN, yes	1.72 (0.98–3.04)	0.06[d]
eGFR per 10 mL/min/1.73 m^2 decrease	1.09 (0.98–1.2)	0.10
NIH LN chronicity index per 1-point increase	**1.34 (1.16–1.54)**	**0.0001**
NIH LN activity index per 1-point increase	**1.13 (1.03–1.23)**	**0.007**
Prolonged glucocorticoids use, yes[c]	**1.86 (1.01–3.44)**	**0.047**
HCQ exposure (status = yes)[e]	**0.36 (0.19–0.68)**	**0.002**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[f]	1.46 (0.71–3.03)	0.30

* eGFR was calculated using the Chronic Kidney Disease Epidemiology Collaboration formular. Variables with a significant *P* value (<0.05) are shown in bold font. ACE-i, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; CI, confidence interval; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; HR, hazard ratio; HTN, hypertension; LN, lupus nephritis; MMF, mycophenolate mofetil; PH, proportional hazards.
[a] Adjusted for age, race, sex, insurance type/status at diagnosis, history of HTN, eGFR at LN diagnosis, NIH LN chronicity and activity indices, mycophenolate duration, ACE-i use, prolonged glucocorticoids use (defined as >7.5 mg of prednisone or equivalent doses for >6 mo), HCQ exposure status, and weight-based HCQ dose categories; insurance and history of HTN were stratified based on Schoenfeld residual analysis to satisfy PH assumption for this model (global PH > 0.05).
[b] Adjusted for age, race, insurance type/status at diagnosis, sex, history of HTN, eGFR at LN diagnosis, NIH LN chronicity and activity indices, prolonged glucocorticoids use, HCQ exposure status, and weight-based HCQ dose categories; race and insurance status/type were stratified based on Schoenfeld residual analysis to satisfy PH assumption (global PH > 0.05).
[c] Defined as glucocorticoids use ≥7.5 mg for >6 mo after LN diagnosis.
[d] Trend toward statistical significance.
[e] HCQ exposure status was abstracted and used as a time-dependent covariate.
[f] HCQ dose >5 mg/kg/d is compared to a combination of HCQ dose ≤5 mg/kg/d or no HCQ use.

including only patients with CKD stage ≥3 at LN diagnosis, HCQ exposure was associated with a 77% lower risk of eGFR decline of ≥30% even in this high-risk group (adjusted HR 0.23 [95% CI 0.06–0.83]; *P* = 0.03; Supplementary Table 4).

Time-varying treatment effect of HCQ on annual eGFR slope decline. Average number of observations per patient were similar in groups with and without HCQ exposure (Figure 1). Adjusted annual eGFR slope decline with time-varying HCQ exposure was –0.65 mL/min/1.73 m^2 (95% CI –0.06 to –1.26 mL/min/1.73 m^2) over 10 years (Figure 1A), whereas the adjusted annual eGFR slope decline without HCQ was five times steeper (–3.82 mL/min/1.73 m^2/year [95% CI –1.20 to –6.44 mL/min/1.73 m^2/year]; Figure 1B) in the first 10 years after diagnosis. A significant reduction in annual eGFR slope decline by 3.17 mL/min/1.73 m^2/year (*P* < 0.0001) was noted with HCQ exposure. A yearly saving of 5.12 mL/min/1.73 m^2 in the adjusted annual eGFR slope decline was noted within 5 years of diagnosis with HCQ (Supplementary Figure 1).

Testing treatment effect of HCQ on eGFR slope using a prevalent LN cohort. *Prevalent LN cohort characteristics.* Key characteristics of the prevalent LN cohort (N = 819) were as follows: mean age 54 ± SD 17 years, 59% women, 83% of White race, 11% of Black race, 4% of Hispanic ethnicity, and 6% having Medicaid. At the first visit, mean eGFR was 67 ± SD 31 mL/min/1.73 m^2, 24% had HTN, 4% had nephrotic syndrome, 71% were ACE-i users, and 32% had been exposed to HCQ.

Treatment effect of HCQ on annual eGFR slope decline in the prevalent cohort. Similar time-varying effects were noted with

Table 3. IPW-adjusted and crude Cox models showing associations between HCQ exposure and HCQ dose with sustained eGFR decline of ≥30% and ≥40%, N = 209*

	Adjusted HR (95% CI)	*P* value
Crude vs IPW-adjusted Cox models for eGFR decline ≥30%		
Crude adjusted Cox model[a]		
HCQ exposure (status = yes)	**0.36 (0.19–0.68)**	**0.0016**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[e]	1.46 (0.71–3.03)	0.30
IPW-adjusted[b] model		
HCQ exposure (status = yes)	**0.40 (0.23–0.69)**	**0.001**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[e]	**2.02 (1.13–3.62)**	**0.02**
Crude vs IPW-adjusted Cox models for eGFR decline ≥40%		
Crude adjusted Cox model[c]		
HCQ exposure (status = yes)	**0.33 (0.17–0.65)**	**0.002**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[e]	1.18 (0.53–2.66)	0.68
IPW-adjusted[d] model		
HCQ exposure (status = yes)	**0.38 (0.21–0.70)**	**0.002**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[e]	1.49 (0.75–2.96)	0.26

* Variables with a significant *P* value (<0.05) are shown in bold font. ACE-i, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; HR, hazard ratio; HTN, hypertension; IPW, inverse probability weighting; LN, lupus nephritis; MMF, mycophenolate mofetil; PH, proportional hazards.
[a] Adjusted for age, race, insurance type/status at diagnosis, sex, history of HTN, eGFR at LN diagnosis, NIH LN chronicity and activity indices, prolonged glucocorticoids use, HCQ exposure status, weight-based HCQ dose categories; race and insurance status/type were stratified based on Schoenfeld residual analysis to satisfy PH assumption (global PH > 0.05).
[b] IPW models adjust for propensity scores for the following covariates: age, sex, race, insurance, history of HTN, baseline eGFR, NIH LN chronicity and activity indices, prolonged glucocorticoid use, MMF duration (in years), LN class, ACE-i/ARB ever use, baseline HCQ use, and median weight-based HCQ dose categories.
[c] Adjusted for age, race, insurance type/status at diagnosis, sex, history of HTN, eGFR at LN diagnosis, NIH LN chronicity and activity indices, HCQ exposure status, and weight-based HCQ dose categories.
[d] IPW models adjust for propensity scores for the following covariates: age, sex, race, insurance, history of HTN, baseline eGFR, NIH LN chronicity and activity indices, prolonged glucocorticoid use, MMF duration (in years), LN class, ACE-i/ARB ever use, baseline HCQ use, and median weight-based HCQ dose categories.
[e] HCQ dose >5 mg/kg/d is compared to a combination of HCQ dose ≤5 mg/kg/d or no HCQ use.

HCQ exposure on annual eGFR slope decline in a separate prevalent LN cohort with longer follow-up. The adjusted annual eGFR slope decline with HCQ was half the annual slope decline noted without HCQ exposure (–0.94 vs 2.23 mL/min/1.73 m^2/year; $P < 0.0001$; Figure 2A–B). Time-varying treatment effect of HCQ on adjusted annual eGFR slope decline was 1.26 mL/min/1.73 m^2 (Figure 2).

DISCUSSION

Kidney involvement is one of the most common severe manifestations of SLE that occurs in up to 70% of patients, and up to 30% of patients experience irreversible kidney damage requiring renal replacement therapy. Literature in SLE highlights lower LN flare risk and prolonged survival with HCQ use, yet information regarding the time-varying treatment effect of HCQ use on preservation of kidney function (eGFR) is limited. Our study leveraged a retrospective biopsy-proven incident LN cohort and reported 59% and 66% lower risk of sustained eGFR decline of ≥30% and ≥40% with time-varying HCQ exposure. Moreover, HCQ exposure was associated with 77% lower risk of eGFR decline of ≥30% in patients with CKD stage ≥3. Using linear mixed-effects modeling to estimate the treatment effect of HCQ on annual eGFR slope, we are the *first* to report a yearly saving in annual eGFR slope decline by 5.12 and 3.17 mL/min/1.73 m^2 with HCQ within 5 and 10 years of diagnosis using our incident LN cohort. A similar reduction in annual eGFR slope decline by 1.26 mL/min/1.73 m^2 was noted with HCQ exposure in a separate large EHR-based prevalent LN cohort with longer follow-up. These findings directly address existing gaps in the literature, reduce uncertainty regarding long-term impact of HCQ use on kidney function, and highlight a benefit of early initiation of HCQ in LN, thereby empowering clinicians to advocate for universal use HCQ in LN.

There are limited data on the longitudinal impact of HCQ on kidney function decline over time.[9,20] Moreover, traditional clinical end points, such as kidney failure or doubling of serum creatinine, require large sample sizes and long follow-up, resulting in insufficient power and low precision in current literature.[5,7,9,34–39] Using surrogate end points such as eGFR slope overcomes these limitations, and eGFR slope analysis can examine the impact of treatment on preservation of kidney function over time. Our study directly addresses this gap by leveraging a linear mixed-effects model to estimate the time-varying treatment effect of HCQ exposure. This analysis revealed that HCQ exposure predicted an attenuation in adjusted annual eGFR slope decline by 3.17 mL/min/1.73 m^2/year ($P < 0.0001$) over 10 years after LN diagnosis in our incident biopsy-proven LN cohort. More importantly, a yearly saving of 5.12 mL/min/1.73 m^2 was noted in eGFR slope decline with HCQ exposure within the first five years of LN diagnosis, indicating early treatment benefits. Finally, using a separate

Table 4. Adjusted Cox regression model showing associations between time-varying HCQ exposure and eGFR decline of ≥40% (N = 209)*

Variables	Adjusted HR (95% CI)[a]	*P* value
Full multivariable Cox model[a]		
Age in 1-y increments	1.01 (0.99–1.03)	0.57
Female	1.2 (0.56–2.59)	0.64
History of HTN, yes	**2.12 (1.09–4.15)**	**0.027**
eGFR per 10 mL/min/1.73 m^2 decrease	1.07 (0.95–1.20)	0.27
NIH LN chronicity index per 1-point increase	**1.33 (1.13–1.57)**	**0.001**
NIH LN activity index per 1-point increase	1.10 (1.00–1.22)	0.053[b]
MMF exposure duration per 1-y increase	0.95 (0.85–1.07)	0.40
ACE-i/ARB ever use, yes	0.77 (0.40–1.51)	0.45
Prolonged glucocorticoids use, yes[c]	1.66 (0.84–3.32)	0.15
HCQ exposure (status = yes)[d]	**0.34 (0.16–0.70)**	**0.004**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[e]	1.55 (0.66–3.65)	0.32
Reduced multivariable Cox model[f]		
Age in 1-y increments	1.00 (0.98–1.03)	0.66
Female	1.21 (0.60–2.41)	0.60
White race	ref	
Asian race	1.45 (0.59–3.57)	0.75
Black race	1.20 (0.40–3.60)	0.42
Hispanic ethnicity	1.09 (0.25–4.83)	0.91
No insurance or Medicaid	**2.68 (1.42–5.06)**	**0.002**
History of HTN, yes	1.76 (0.94–3.28)	0.08[d]
eGFR per 10 mL/min/1.73 m^2 decrease	1.07 (0.96–1.20)	0.23
NIH LN chronicity index per 1-point increase	**1.30 (1.12–1.52)**	**0.001**
NIH LN activity index per 1-point increase	**1.12 (1.01–1.23)**	**0.03**
HCQ exposure (status = yes)[d]	**0.33 (0.17–0.65)**	**0.002**
Higher than guideline-recommended HCQ dose >5 mg/kg/d, yes[e]	1.18 (0.53–2.66)	0.68

* eGFR calculated using the Chronic Kidney Disease Epidemiology Collaboration formular. Variables with a significant *P* value (<0.05) are shown in bold font. ACE-i, angiotensin converting enzyme inhibitor; ARB, angiotensin receptor blocker; CI, confidence interval; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; HR, hazard ratio; HTN, hypertension; LN, lupus nephritis; MMF, mycophenolate mofetil; PH, proportional hazards.
[a] Adjusted for age, race, sex, insurance type/status at diagnosis, history of HTN, eGFR at LN diagnosis, NIH LN chronicity and activity indices, mycophenolate duration, ACE-i use, prolonged glucocorticoids use (defined as >7.5 mg of prednisone or equivalent doses for >6 mo), HCQ exposure status, and weight-based HCQ dose categories; race and insurance were stratified based on Schoenfeld residual analysis to satisfy PH assumption for this model (global PH > 0.05).
[b] Indicates trend toward significance.
[c] Defined as glucocorticoids use ≥7.5 mg for >6 mo after LN diagnosis.
[d] HCQ exposure status was abstracted and used as a time-dependent covariate.
[e] HCQ dose >5 mg/kg/d is compared to a combination of HCQ dose ≤5 mg/kg/d or no HCQ use
[f] Adjusted for age, race, insurance type/status at diagnosis, sex, history of HTN, eGFR at LN diagnosis, NIH LN chronicity and activity indices, HCQ exposure status, weight-based HCQ dose categories, and PH assumption (global PH > 0.05).

prevalent LN cohort with longer follow-up, HCQ use again was associated with an attenuation in annual eGFR slope by 1.26 mL/min/1.73 m^2 over a longer follow-up period. A meta-regression found that an attenuation in the rate of decline in eGFR of ≥0.75 mL/min/1.73 m^2 per year can reduce the risk of kidney failure progression by 16% to 23%.[5,38] Thus, our observed attenuation in eGFR decline over different time periods (range = 1.26–5.12 mL/min/1.73 m^2) would be associated with at least 23%, if not more, reduced kidney failure risk. Moreover as an adjunct therapy for LN, HCQ could enhance effectiveness of LN induction and maintenance therapies.[5] Thus, our study uniquely highlights the critical impact of HCQ use on preserving kidney function and underscores the need to use HCQ early and for a longer duration in LN.

Existing literature shows an overall positive association between HCQ use and preventing renal damage.[11,13–16] Particularly, Pons-Estel et al noted a 70% reduction in the risk of renal damage (defined as 50% reduction in eGFR) with HCQ use.[34] Likewise, recent studies using cohorts from Israel and Mexico, reported 0.4-fold lower risk of CKD progression to stage ≥3 or kidney failure in HCQ users.[14,16] Despite strong associations and universal recommendation to use HCQ in LN, <70% of patients get started on HCQ.[11,12,40] Underuse of HCQ could be linked to current gaps in literature: (1) low precision in risk estimates (wide CIs) given few terminal events (kidney failure or doubling of serum creatinine or CKD progression) occurred over time,[7] (2) contrasting evidence when time was used as a covariable highlighting no association or worsening of podocyte function with HCQ in LN possibly due to immortal time and survival biases,[10,41] (3) timing and duration of HCQ therapy is not clear, and (4) treatment effect of HCQ in CKD stage ≥3 needs clarification.[9,40] Such gaps can exacerbate existing fears to use HCQ,

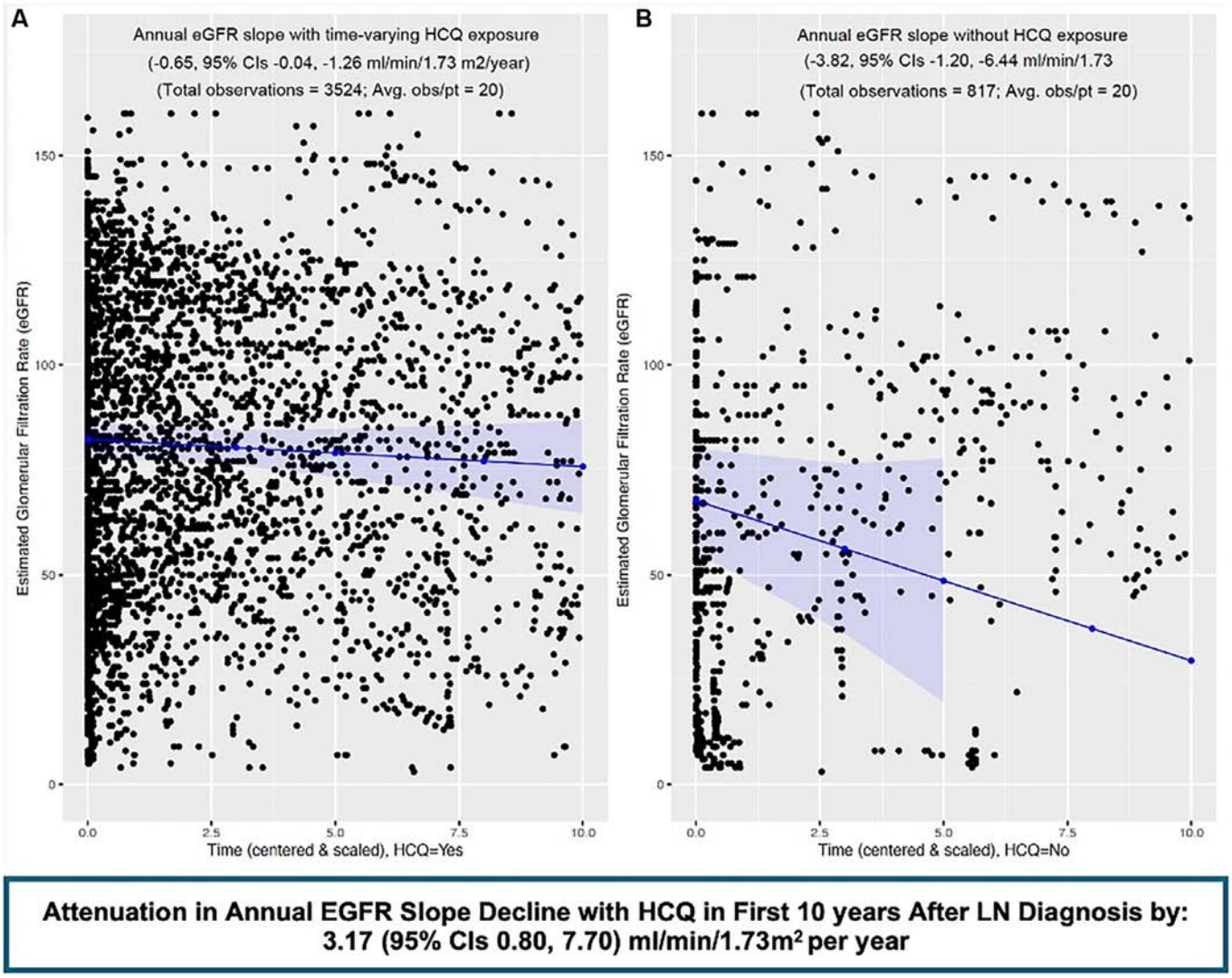


Figure 1. (A) Annual eGFR slope decline with HCQ use versus (B) without HCQ use, using random slope and intercept linear mixed-methods model adjusting for covariables (age, race, sex, LN pathology findings, comorbidities, treatment) in our LN inception cohort (N = 209). CI, confidence interval; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; LN, lupus nephritis.

particularly in nonrheumatology clinicians.[40] Our study directly addresses these four gaps and bolsters the evidence regarding HCQ's specific benefit in LN. First, by using sustained eGFR ≥30% or ≥40% and eGFR slope decline as surrogate end points of kidney failure, we were able to improve precision of risk estimates. Second, even after propensity score adjustments and using HCQ exposure as a time-dependent covariate, a significant 54% and 51% reduction in end points, eGFR decline of ≥30% and ≥40%, was noted with HCQ exposure. Third, a significant attenuation in eGFR slope and a 73% lower risk of eGFR decline of ≥30% was noted with early HCQ use, within five years of diagnosis. Finally, our subgroup analysis highlighted a 77% lower risk of eGFR decline of ≥30% with HCQ use in patients with CKD stage ≥3. By addressing the concerns about why, when, and how long HCQ should be used in LN, including those with CKD, our findings can support clinicians and patients in shared decision-making, encouraging higher uptake of HCQ.

This study has several strengths including using the following: (1) a validated incident biopsy-proven LN cohort, (2) HCQ exposure as a time-dependent covariate, and (3) a robust linear mixed-effects analysis to determine treatment effect of HCQ on eGFR slope decline. However, we acknowledge limitations. First, this study uses data from a single academic institution and may not represent the LN population in the US. Second, we did not have data on HCQ blood levels to account for variability in adherence and absorption of HCQ. Third, we could have missed eGFR decline in patients who moved, lost follow-up, or died in outside institutions; however, several sensitivity analyses were performed. Fourth, given the retrospective design of the study, we could not account for unknown confounders and establish causal associations between HCQ exposure and kidney function decline.

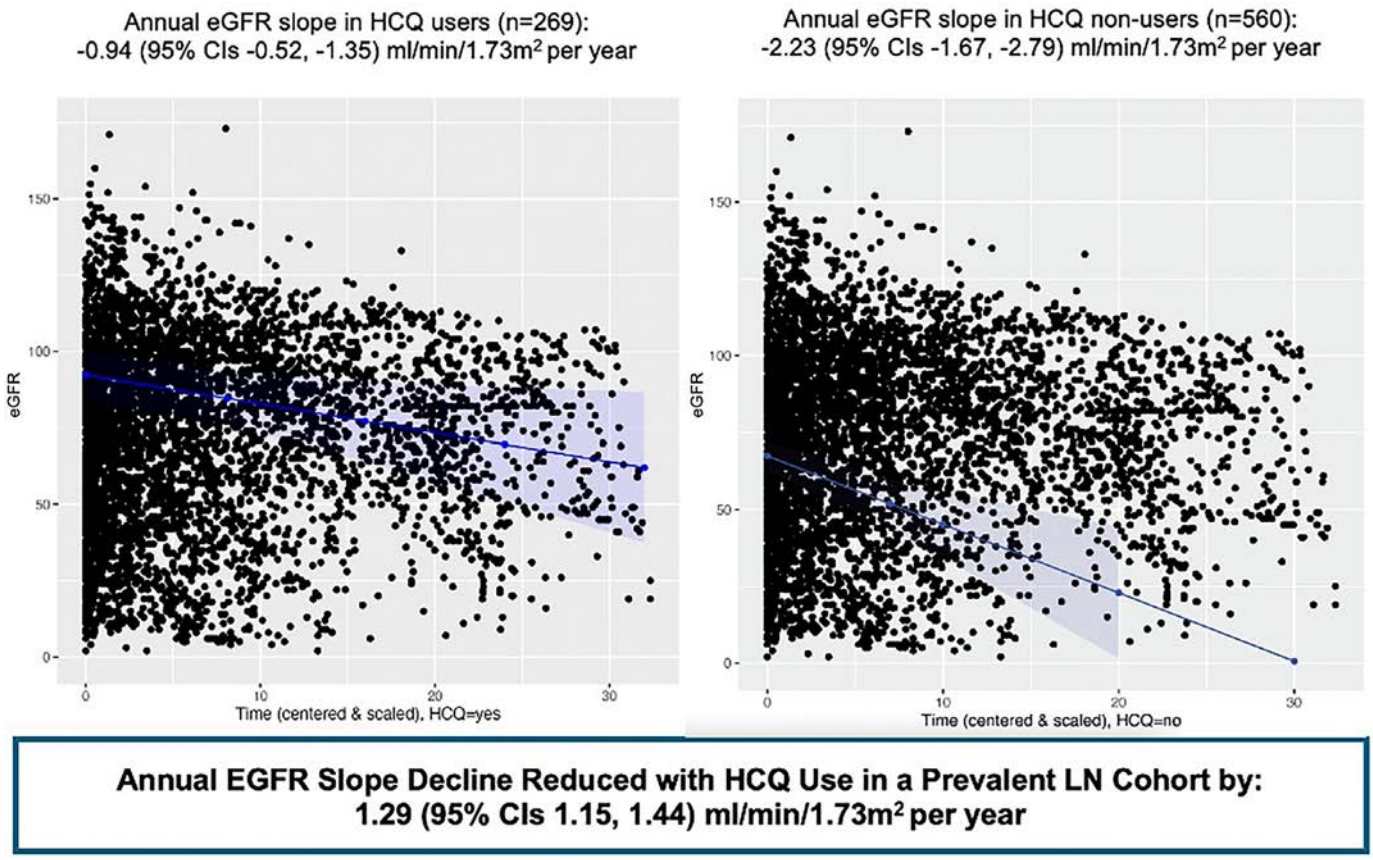


Figure 2. Linear mixed-effects model examining the treatment effect of HCQ on eGFR in a prevalent LN cohort (N = 819 patients, including all visits of patients with ≥2 ICD-10 LN diagnosis codes controlling for confounders [sociodemographics, comorbidities, LN characteristics, ACE-i exposure]). All data points shown in both graphs. CI, confidence interval; eGFR, estimated glomerular filtration rate; HCQ, hydroxychloroquine; ICD-10, International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; LN, lupus nephritis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25616/abstract.

However, we calculated *E* values of 5 and 6, indicating that any unmeasured confounder needs to be associated with five-fold and six-fold higher eGFR decline of ≥30% and ≥40% for the association with HCQ exposure duration to be truly null. Additionally, we performed IPW analysis and Cox analysis using HCQ exposure as a time-dependent covariate to address survival and immortal time bias. Fifth, our CKD stage ≥3 subgroup was small, and our findings need to be validated in larger cohorts. Finally, we could not ascertain the reasons for stopping HCQ or delayed onset of HCQ due to limitations of the study design.

The implications of our findings are three-fold. First, they highlight kidney-specific effects of HCQ exposure duration in LN, particularly on eGFR decline ≥30% or ≥40%, and reduction in annual eGFR slope decline (early surrogate end points of terminal events like kidney failure). Second, our study supports the inclusion of HCQ as a pivotal component of LN management protocols, not just for its systemic benefits, but also for a potential to preserve kidney function. Finally, these findings highlight both early and long-term benefits of using HCQ in attenuating eGFR slope decline in LN. These findings emphasize the importance of starting early and continuing HCQ therapy to preserve kidney health alongside overall disease control even if SLE is limited to the kidneys (LN).

ACKNOWLEDGMENTS

We would like to thank the Office of Informatics and Information Technology at the University of Wisconsin School of Medicine and Public Health for building algorithms for electronic health records data abstraction for incident and prevalent cohorts. We would like to thank Dr Zhong and Dr Panzer for supporting the pathology review.

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 Garg 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. Hocaoğlu M, Valenzuela-Almada MO, Dabit JY, et al. Incidence, prevalence, and mortality of lupus nephritis: a population-based study over four decades using the Lupus Midwest Network. Arthritis Rheumatol 2023;75(4):567–573.
2. Choi HS, Han KD, Jung JH, et al. The risk of end-stage renal disease in systemic lupus erythematosus: a nationwide population-based study in Korea. Medicine (Baltimore) 2019;98(28):e16420.
3. Wasik H, Chadha V, Galbiati S, et al. Dialysis outcomes for children with lupus nephritis compared to children with other forms of nephritis: a retrospective cohort study. Am J Kidney Dis 2022;79(5): 626–634.
4. Wong T, Goral S. Lupus nephritis and kidney transplantation: where are we today? Adv Chronic Kidney Dis 2019;26(5):313–322.
5. Rovin BH, Furie RA, Ross Terres JA, et al. Kidney outcomes and preservation of kidney function with obinutuzumab in patients with lupus nephritis: a post hoc analysis of the NOBILITY trial. Arthritis Rheumatol 2024;76(2):247–254.
6. Hahn BH, McMahon MA, Wilkinson A, et al; American College of Rheumatology. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. Arthritis Care Res (Hoboken) 2012;64(6):797–808.
7. Rovin BH, Ayoub IM, Chan TM, et al. Executive summary of the KDIGO 2024 Clinical Practice Guideline for the Management of Lupus Nephritis. Kidney Int 2024;105(1):31–34.
8. Sammaritano LR, Askanase A, Bermas BL, et al. 2024 American College of Rheumatology (ACR) Guideline for the Screening, Treatment, and Management of Lupus Nephritis. Arthritis Rheumatol 2025; 77(9):1115–1135.
9. Wu CY, Tan M, Huang JY, et al. Hydroxychloroquine is neutral in risk of chronic kidney disease in patients with systemic lupus erythematosus. Ann Rheum Dis 2022;81(5):e75.
10. An N, Yang C, Wu HL, et al. Hydroxychloroquine administration exacerbates acute kidney injury complicated by lupus nephritis. Arthritis Res Ther 2022;24(1):6.
11. Xiong WW, Boone JB, Wheless L, et al. Real-world electronic health record identifies antimalarial underprescribing in patients with lupus nephritis. Lupus 2019;28(8):977–985.
12. Garg S, Singh T, Panzer SE, et al. Multidisciplinary lupus nephritis clinic reduces time to renal biopsy and improves care quality. ACR Open Rheumatol 2022;4(7):581–586.
13. Pons-Estel GJ, Alarcón GS, McGwin G Jr, et al; Lumina Study Group. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. Arthritis Rheum 2009;61(6):830–839.
14. Pokroy-Shapira E, Gelernter I, Molad Y. Evolution of chronic kidney disease in patients with systemic lupus erythematosus over a long-period follow-up: a single-center inception cohort study. Clin Rheumatol 2014;33(5):649–657.
15. Gheet FS, Dawoud HE, El-Shahaby WA, et al. Hydroxychloroquine in children with proliferative lupus nephritis: a randomized clinical trial. Eur J Pediatr 2023;182(4):1685–1695.
16. Peña-Vizcarra ÓR, Zavala-Miranda MF, Juárez-Cuevas B, et al. Effect of antimalarials on clinical outcomes in lupus nephritis. Rheumatology (Oxford) 2024:63(8):2230–2238.
17. Sisó A, Ramos-Casals M, Bové A, et al. Previous antimalarial therapy in patients diagnosed with lupus nephritis: influence on outcomes and survival. Lupus 2008;17(4):281–288.
18. Levin A, Agarwal R, Herrington WG, et al; participant authors of the International Society of Nephrology's 1st International Consensus Meeting on Defining Kidney Failure in Clinical Trials. International consensus definitions of clinical trial outcomes for kidney failure: 2020. Kidney Int 2020;98(4):849–859.
19. Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. Am J Kidney Dis 2014;64(6):821–835.
20. Vinet E, Bernatsky S, Suissa S. Have some beneficial effects of hydroxychloroquine been overestimated? Potential biases in observational studies of drug effects: comment on the article by Pons-Estel et al. Arthritis Rheum 2009;61(11):1614–1615.
21. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40(9):1725.
22. Petri M, Orbai AM, Alarcón GS, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64(8): 2677–2686.
23. Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology Classification Criteria for Systemic Lupus Erythematosus. Arthritis Rheumatol 2019;71(9):1400–1412.
24. Weening JJ, D'Agati VD, Schwartz MM, et al; International Society of Nephrology Working Group on the Classification of Lupus Nephritis; Renal Pathology Society Working Group on the Classification of Lupus Nephritis. The classification of glomerulonephritis in systemic lupus erythematosus revisited. Kidney Int 2004;65(2):521–530.
25. Kind AJH, Jencks S, Brock J, et al. Neighborhood socioeconomic disadvantage and 30-day rehospitalization: a retrospective cohort study. Ann Intern Med 2014;161(11):765–774.
26. Lloyd-Jones DM, Braun LT, Ndumele CE, et al. Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: a special report from the American Heart Association and American College of Cardiology. J Am Coll Cardiol 2019;73(24):3153–3167.
27. Liapis H, Gaut JP, Klein C, et al; Banff Working Group. Banff Histopathological Consensus Criteria for Preimplantation Kidney Biopsies. Am J Transplant 2017;17(1):140–150.
28. Zhang Z, Reinikainen J, Adeleke KA, et al. Time-varying covariates and coefficients in Cox regression models. Ann Transl Med 2018: 6(7):121.
29. Fanouriakis A, Kostopoulou M, Alunno A, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. Ann Rheum Dis 2019;78(6):736–745.
30. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.
31. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Ann Intern Med 2017;167(4): 268–274.
32. Izadi Z, Gianfrancesco M, Anastasiou C, et al. Development and validation of a risk scoring system to identify patients with lupus nephritis in electronic health record data. Lupus Sci Med 2024;11(1):e001170.
33. Murphy JI, Weaver NE, Hendricks AE. Accessible analysis of longitudinal data with linear mixed effects models. Dis Model Mech 2022; 15(5):dmm048025.
34. Pons-Estel GJ, Alarcón GS, Hachuel L, et al; GLADEL. Anti-malarials exert a protective effect while Mestizo patients are at increased risk of developing SLE renal disease: data from a Latin-American cohort. Rheumatology (Oxford) 2012;51(7):1293–1298.
35. Pons-Estel GJ, González LA, Zhang J, et al. Predictors of cardiovascular damage in patients with systemic lupus erythematosus: data from LUMINA (LXVIII), a multiethnic US cohort. Rheumatology (Oxford) 2009;48(7):817–822.

36. Fanouriakis A, Tziolos N, Bertsias G, et al. Update on the diagnosis and management of systemic lupus erythematosus. Ann Rheum Dis 2021;80(1):14–25.
37. Fanouriakis A, Kostopoulou M, Cheema K, et al. 2019 Update of the Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of lupus nephritis. Ann Rheum Dis. 2020;79(6):713–723.
38. Odler B, Fu EL. eGFR slope as a primary endpoint for clinical trials of CKD progression: one size fits all? Clin Kidney J 2024;17(1):sfae001.
39. Bertsias GK, Tektonidou M, Amoura Z, et al; European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association. Joint European League Against Rheumatism and European Renal Association-European Dialysis and Transplant Association (EULAR/ERA-EDTA) recommendations for the management of adult and paediatric lupus nephritis. Ann Rheum Dis 2012;71(11):1771–1782.
40. Ayoub I, Singh P, Ardoin S, et al. What every nephrologist needs to know about hydroxychloroquine toxicity. Clin Nephrol 2020;93(3): 149–151.
41. Wu T, Huang W, Qi J, et al. Research trends and frontiers on antiphospholipid syndrome: a 10-year bibliometric analysis (2012-2021). Front Pharmacol 2022;13:1035229.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 249–258
DOI 10.1002/acr.25620

American College of Rheumatology
Empowering Rheumatology Professionals

Heterogeneity of Rheumatoid Arthritis–Associated Interstitial Lung Disease by Longitudinal Forced Vital Capacity Trajectory and Associations With Disease Outcomes

Bryant R. England,[1] Yangyuna Yang,[1] Punyasha Roul,[1] Tate M. Johnson,[1] Scott M. Matson,[2] Grant W. Cannon,[3] Brian C. Sauer,[3] Jeffrey R. Curtis,[4] Joshua F. Baker,[5] and Ted R. Mikuls[1]

Objective. We aimed to identify unique disease trajectories within rheumatoid arthritis–associated interstitial lung disease (RA-ILD) based on longitudinal forced vital capacity (FVC) values and their associated clinical outcomes.

Methods. We performed a cohort study of RA-ILD within the Veterans Health Administration from 1999 to 2021. Patients with RA-ILD were identified with validated algorithms. Group-based trajectory modeling was used to identify unique groups based on longitudinal FVC values in a primary cohort with baseline and follow-up FVC values and an overall cohort with at least one FVC value. Associations of group assignment with survival and respiratory hospitalization were tested in multivariable Cox regression models adjusting for initial FVC.

Results. We derived three FVC trajectory groups in the primary cohort (n = 1,092) and eight in the overall cohort (n = 5,172). A total of 82.5% (primary cohort) and 54% (overall cohort) of patients with RA-ILD belonged to progressive groups with FVC percent predicted change ranging from –2.16 to –10.78 (primary) and –0.73 to –1.56 (overall) per year. Relative to the stable group, slow (adjusted hazard ratio [aHR] 1.55 and 1.38) and rapid (aHR 2.07 and 1.77) progressors in the primary cohort had an increased risk of death and respiratory hospitalization, respectively. In the overall cohort, patients with a progressive FVC trajectory also experienced an increased risk of death (aHR range 1.18–2.69) and respiratory hospitalization (aHR range 1.29–2.78).

Conclusion. Unique disease trajectories based on longitudinal FVC values are independently associated with the risk of death and respiratory-related hospitalization in RA-ILD. Recognition of these unique RA-ILD trajectories could inform management and monitoring.

INTRODUCTION

Interstitial lung disease (ILD) is an inflammatory and fibrotic diffuse lung disease that affects between 10% and 40% of people with rheumatoid arthritis (RA).[1] Individuals afflicted with RA-ILD suffer from reduced quality of life, poor survival, and a high degree of health care use.[2–4] Recognizing the challenge providers face managing RA-ILD in the absence of high quality evidence, the

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government.

Supported by Boehringer-Ingelheim and the Center of Excellence for Suicide Prevention, Joint Department of Veterans Affairs (VA) and Department of Defense Mortality Data Repository – National Death Index. Support for VA/Centers for Medicare and Medicaid Services data was provided by the Department of Veterans Affairs, VA Health Services Research and Development Services, and VA Information Resource Center (SDR 02-237 and 98-004). Dr England's work was supported by the VA Clinical Science Research and Development (CSR&D; grant IK2 CX002203) and Rheumatology Research Foundation. Dr Johnson's work was supported by the Rheumatology Research Foundation. Dr Matson's work was supported by the National Institutes of General Medical Sciences, NIH (grant P20 GM130423). Dr Curtis's work was supported by the National Institutes of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant P30 AR072583). Dr Baker's work was supported by the CSR&D Merit Award (grant CX001703) and Rehabilitation Research and Development Merit Award (grant RX003644). Dr Mikuls's work was supported by the VA Biomedical Laboratory Research and Development Merit Award (grant I01 BX003635), NIH (grant 2U54GM115458), US Department of Defense (grant PR200793), and Rheumatology Research Foundation.

[1]Bryant R. England, MD, PhD, Yangyuna Yang, MBBS, PhD, Punyasha Roul, MS, Tate M. Johnson, MD, Ted R. Mikuls, MD, MSPH: VA Nebraska-Western Iowa Health Care System and University of Nebraska Medical Center, Omaha; [2]Scott M. Matson, MD: Kansas University Medical Center, Kansas City; [3]Grant W. Cannon, MD, Brian C. Sauer, PhD: Salt Lake City VA Health Care System and University of Utah, Salt Lake City; [4]Jeffrey R. Curtis, MD, MS, MPH: University of Alabama at Birmingham; [5]Joshua F. Baker, MD, MSCE: Corporal Michael J. Crescenz VA Medical Center and University of Pennsylvania, Philadelphia.

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.25620).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25620.

Address correspondence via email to Bryant R. England, MD, PhD, at Bryant.england@unmc.edu.

Submitted for publication June 16, 2025; accepted in revised form July 22, 2025.

SIGNIFICANCE & INNOVATIONS

- Rheumatoid arthritis–associated interstitial lung disease (RA-ILD) is a heterogeneous disorder with recognized variability by ILD pattern, change in pulmonary function (eg, forced vital capacity [FVC]), and disease outcomes.
- Using group-based trajectory modeling of a large RA-ILD cohort, we identified slowly progressive, rapidly progressive, and stable FVC disease trajectory groups, with the majority of patients with RA-ILD belonging to progressive FVC decline groups.
- Patients characterized by slow or rapid FVC progression experienced an increased risk of death and respiratory-related hospitalization, despite the rate of FVC decline being below thresholds used for eligibility criteria in antifibrotic clinical trials.
- Our findings suggest that reliance on the progression thresholds used as part of therapeutic regulatory trials in ILD may be overly stringent to inform real-world treatment decisions in RA-ILD.

American College of Rheumatology (ACR) and American College of Chest Physicians (CHEST) recently developed the first clinical practice guidelines for the treatment of systemic autoimmune and rheumatic disease–associated ILD, including RA-ILD.[5] In that guideline, all treatment recommendations for RA-ILD were conditional in nature because of the limited evidence and the phenotypic variability in RA-ILD among patients. This substantial heterogeneity occurring in patients with RA-ILD complicates disease management.

RA-ILD is commonly categorized according to the historadiologic ILD pattern. Usual interstitial pneumonia (UIP) is the most common pattern in RA-ILD, with nonspecific interstitial pneumonia being the most common non-UIP ILD pattern.[6] Although genetic risk factors (eg, *MUC5B* promoter variant) differ by ILD pattern, and patients with RA-ILD with a UIP pattern appear to have a higher risk of death than those with other patterns,[7–9] ILD pattern was not considered a factor influencing treatment recommendations in the ACR/CHEST guidelines.[5] Supporting this, an observational study recently demonstrated the ability of azathioprine, mycophenolate mofetil, and rituximab to all stabilize forced vital capacity (FVC), findings that did not differ by ILD pattern.[10] Similarly, post hoc analyses of the INBUILD trial found similar slowing of FVC decline with nintedanib compared with placebo among patients with progressive fibrosing ILD, regardless of ILD pattern.[11] Thus, other indicators of disease heterogeneity besides ILD pattern on chest computed tomography (CT) are needed to guide RA-ILD management.

Although the management implications of ILD pattern in RA-ILD remain to be fully elucidated, the importance of considering disease progression by pulmonary function test (PFT) parameters is becoming more apparent. In fact, based on the aforementioned INBUILD trial,[11] nintedanib received United States Food and Drug Administration approval for ILD specifically with a progressive fibrotic phenotype. Recognizing the importance of characterizing longitudinal pulmonary function in RA-ILD, recent studies have preliminarily explored whether there are underlying differences in the clinical disease course within RA-ILD.[12,13] Two studies identified trajectories corresponding to improvement and/or stability, slow progression, and rapid progression that are predictive of disease outcomes (eg, death), though these reports were limited by small sample sizes (n = 140–172 RA-ILD patients) and single-center study populations.[12,13] Thus, our understanding of the range of different RA-ILD disease trajectories that occur in large, real-world RA-ILD populations remains poorly understood.

The purpose of this study was to identify unique trajectories of FVC change over time among a national-scale, real-world population of patients with RA-ILD within the Veterans Health Administration (VA), the largest integrated health care system in the United States. Subsequently, we aimed to evaluate the prognosis of patients within these unique RA-ILD trajectories. We hypothesized that there would be distinct trajectories of FVC among patients with RA-ILD that corresponded to differential prognosis.

PATIENTS AND METHODS

Design and population. We performed a cohort study of patients with RA-ILD in the VA between 1999 and 2021. Patients with RA-ILD were identified based on fulfilling validated algorithms for RA and ILD. For RA, we required the presence of two or more RA diagnostic codes, rheumatologist diagnosis of RA, and receipt of a disease-modifying antirheumatic drug (DMARD) or positive autoantibody (rheumatoid factor [RF] or anti–cyclic citrullinated peptide [anti-CCP] antibody).[14–16] For ILD, we required the presence of two or more ILD diagnostic codes and either a pulmonologist diagnosis of ILD, completion of chest CT and PFTs, or completion of lung biopsy. Individuals with diagnostic codes for other connective tissue diseases (eg, systemic lupus erythematosus, inflammatory myositis, systemic sclerosis) or pneumoconioses were excluded. Such algorithms have >90% positive predictive value (PPV) for RA and >80% to 85% PPV for RA-ILD.[17,18] The study index date was the date patients fulfilled the RA-ILD algorithm (ie, latest of required components), which was at least two years after their first VA health care encounter for 88.2% of patients. For our primary RA-ILD cohort, we required an FVC value within three years before the index date and at least one FVC value within five years after the index date. Because the disease course often impacts FVC testing in real-world settings, which could introduce selection bias, we also created a broader overall RA-ILD cohort that only required patients to have at least one FVC value during the study period. This study received institutional review board (IRB) approval (VA central IRB #1619487 and University of Utah IRB #12917). Patients and the public were not involved in this study.

FVC values. In the primary cohort, FVC values were evaluated within three years before the index date to five years after the index date. In the overall cohort, all available FVC values were included. FVC values were obtained from two sources. First, when PFT equipment was compatible with discrete data storage, FVC values were obtained from the VA Corporate Data Warehouse (CDW). Second, recognizing frequent PFT equipment incompatibility with the VA CDW, we extracted additional FVC values from electronic health record notes using a validated natural language processing (NLP) tool.[19] All FVC values were analyzed in this study as FVC percent predicted values.

Survival and respiratory hospitalization. In cohort analyses evaluating the association of unique disease trajectories with disease outcomes, the study outcomes were all-cause mortality and respiratory-related hospitalization. Vital status was obtained through linkage with the National Death Index, a highly valid approach for ascertaining vital status in veterans.[20] Respiratory-related hospitalizations were obtained from the VA CDW, non-VA care billed to the VA (eg, Fee Basis), and linked Centers for Medicare and Medicaid Services (CMS) data. Hospitalizations were considered respiratory related if the primary discharge diagnosis had respiratory-related International Classification of Diseases, Ninth Revision, (ICD-9; 466.x–519.x) or ICD-10 (J09.x–J99.x) diagnostic codes. These events were assessed after the index date.

Clinical variables. Additional clinical data were obtained for descriptive purposes and to serve as covariables in regression models. These included demographics, smoking status, body mass index (BMI), comorbidity burden, RA autoantibody status (RF and anti-CCP antibody), elevated erythrocyte sedimentation rate or C-reactive protein, DMARD use, and antifibrotic use (nintedanib and pirfenidone). These variables were obtained from the VA CDW before or on the index date. Comorbidity burden was assessed using a count of up to 43 chronic conditions, each requiring multiple diagnostic codes on separate days before the index date, as in prior work.[21] Medication use was evaluated before the index date. We categorized patients into calendar year periods based on the year fulfilling the RA-ILD algorithm (1999–2005, 2006–2011, 2012–2021).

Statistical analysis. Group-based trajectory modeling was used to define unique clusters of patients based on longitudinal FVC trajectory. Group-based trajectory modeling is a finite mixture statistical modeling approach to approximate the unknown distribution of trajectories within a population.[22] These approaches have been previously used in RA-ILD and systemic sclerosis–associated ILD.[12,13,23] Group-based trajectory modeling was implemented using the "traj" Stata plugin (StataCorp). Consistent with best practices for group-based trajectory modeling,[22] the selection of optimal numbers of groups was informed by model fit characteristics (eg, change in Akaike information criterion) and the size of the smallest group combined with clinical interpretability. Linear models were used because polynomial functions were tested and did not improve model fit (data not shown). In the primary RA-ILD cohort, change in FVC from the initial value was modeled. In the broader overall RA-ILD cohort, the actual FVC value was modeled, producing both an intercept (ie, initial FVC value) and slope (ie, change in FVC values). For individuals in the overall cohort with only one FVC value, the proximity of this value to the estimated trajectories determined group assignment.

Patient characteristics were descriptively summarized overall and by group assignment. Associations of FVC trajectory group assignment with survival and respiratory-related hospitalization were assessed using unadjusted Kaplan–Meier curves and in separate multivariable Cox regression models. Covariates in these models were age, sex, race, smoking status, BMI category, calendar year period, comorbidity burden, and baseline FVC. The proportional hazards assumption was tested by Schoenfeld residuals and negative log-log survival plots. Schoenfeld residuals were significant, which was identified to be related to missing or unknown race data, a covariate in our regression models. No other Schoenfeld residuals were significant and log-log survival plots did not indicate violation. Therefore, Cox models were considered appropriate for our analyses. Missing covariables were addressed using the missing indicator method. All analyses were completed using Stata version 18 within the VA Informatics and Computing Infrastructure.

RESULTS

Study population. We studied 5,172 patients with RA-ILD who had 18,007 available FVC values in the overall RA-ILD cohort, of which 1,092 patients with RA-ILD (with 4,038 FVC values) fulfilled additional eligibility criteria for the primary cohort. Patients in both cohorts were predominantly male, had a mean age of 68 years, and had a frequent-smoking history (Table 1). Comorbidity burden was high with a mean of 11.4 (SD 5.3) comorbidities and 49.7% having chronic obstructive pulmonary disease (COPD). Prednisone (67.8% and 72.4%) and conventional synthetic DMARDs (csDMARDs; 64.8 and 69.1%) were frequently used in the primary and overall cohorts, whereas biologic/targeted-synthetic DMARDs were used less frequently (28.7 and 32.9%). Baseline antifibrotic use in the RA-ILD population was rare during this study period (0.2%).

FVC trajectories. *Primary RA-ILD cohort.* We identified three FVC trajectory groups (Supplemental Table 1) that were labeled as stable, slow progression, and rapid progression based on change in FVC over time (Figure 1). Only 17.5% of patients with RA-ILD had a stable FVC trajectory. Most patients were

Table 1. Patient characteristics of study cohorts*

	Primary RA-ILD cohort (n = 1,092)	Overall RA-ILD cohort (n = 5,172)
Age, y	68.0 (9.5)	68.8 (9.9)
Male	1,012 (92.7%)	4,776 (92.3%)
Race		
Asian/multiple races	24 (2.2%)	122 (2.4%)
Black	162 (14.8%)	708 (13.7%)
White	782 (71.6%)	3,882 (75.1%)
Missing	124 (11.4%)	460 (8.9%)
Calendar year		
1999–2005	284 (26.0%)	943 (18.2%)
2006–2011	256 (23.4%)	1,382 (26.7%)
2012–2021	552 (50.5%)	2,847 (55.0%)
Smoking status		
Never	133 (12.2%)	606 (11.7%)
Former	346 (31.7%)	1,619 (31.3%)
Current	463 (42.4%)	2,331 (45.1%)
Missing	150 (13.7%)	616 (11.9%)
Comorbidity burden	10.1 (4.8)	11.4 (5.3)
COPD	481 (44.0%)	2,569 (49.7%)
RF/CCP antibody status		
Seronegative	190 (17.4%)	903 (17.5%)
Seropositive	747 (68.4%)	3,479 (67.3%)
Missing	155 (14.2%)	790 (15.3%)
ESR/CRP		
Normal	230 (21.1%)	1,143 (22.1%)
High	696 (63.7%)	3,402 (65.8%)
Missing	166 (15.2%)	627 (12.1%)
Prednisone[a]	740 (67.8%)	3,744 (72.4%)
csDMARDs[a]	708 (64.8%)	3,573 (69.1%)
b/tsDMARDs[a]	313 (28.7%)	1,702 (32.9%)
Antifibrotic use[a]	2 (0.2%)	9 (0.2%)

* Values are represented as mean (SD) or n (%). b/tsDMARD, biologic/targeted-synthetic disease-modifying antirheumatic drug; CCP, cyclic citrullinated peptide; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; csDMARD, conventional synthetic disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; RA-ILD, rheumatoid arthritis–associated interstitial lung disease; RF, rheumatoid factor.
[a] Use before the study index date.

categorized into the slow progression group (75.8%), with a mean FVC change of –2.16% per year. Rapid progression occurred among 6.7%, with a mean FVC change of –10.78% per year. Those with progressive trajectories were older, more likely men and former smokers, and less likely to have comorbid COPD (Supplemental Table 2).

Overall RA-ILD cohort. In the overall cohort, eight unique FVC trajectory groups were identified (Supplemental Table 3) that varied both by the initial FVC value (ie, intercept) and rate of FVC change during follow-up (Figure 1). Of the eight RA-ILD trajectory groups, we categorized four groups as progressive based on their observed rate of change. The mean decline in FVC percent predicted per year ranged from –0.73% to –1.56% in the progressive group. A total of 54% of the patients with RA-ILD belonged to one of the four progressive RA-ILD trajectory groups. The intercept for initial FVC percent predicted also varied among these progressive ILD patients, ranging from 40.9% to 81.1%. Among the 46% of patients with RA-ILD belonging to the four nonprogressive groups, mean change in FVC percent predicted per year ranged from –0.29% to 0.36%. In these groups, the intercept for initial FVC ranged from 63.9% to 110.3%.

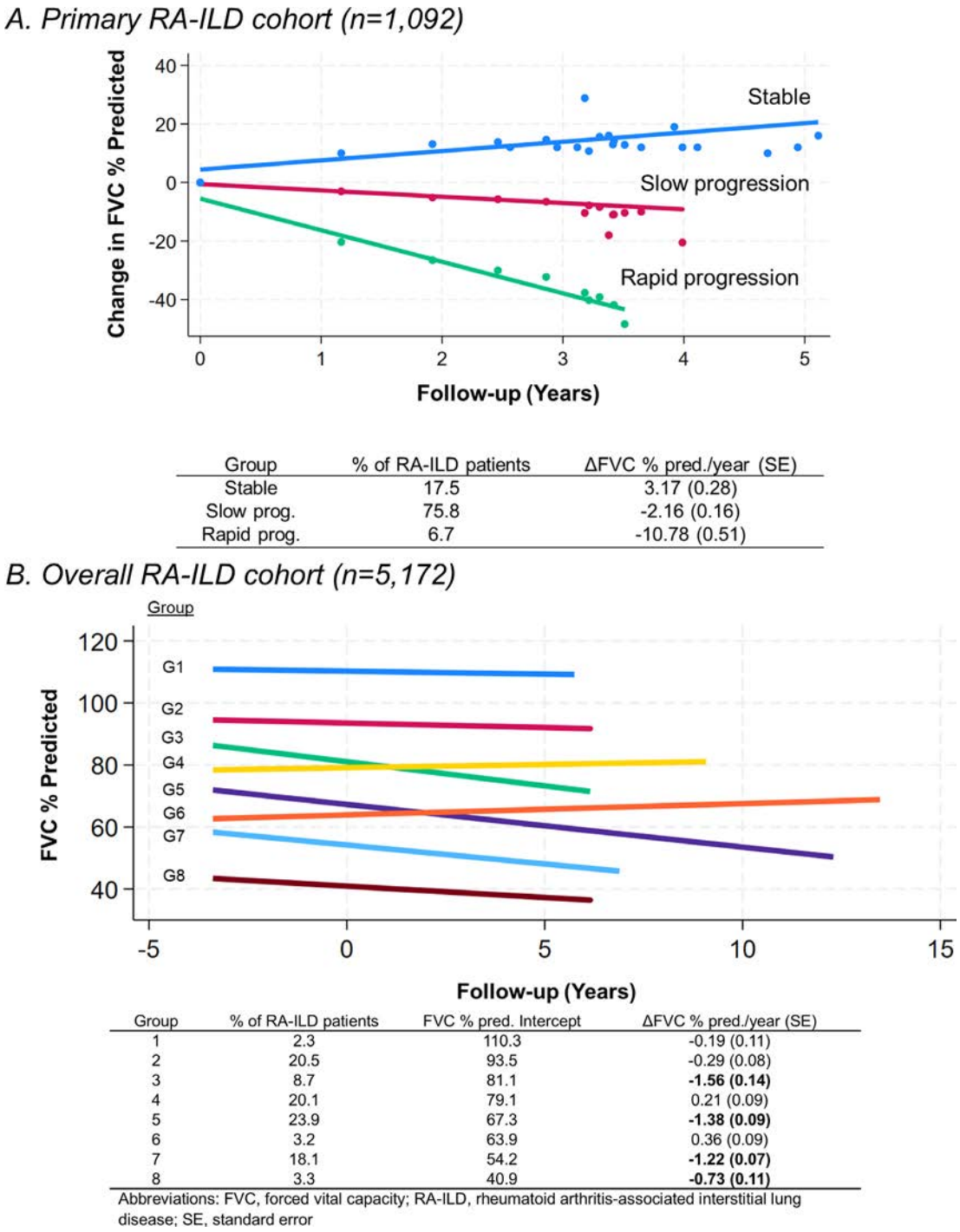
Those with a progressive FVC trajectory were more likely to have been followed from earlier calendar year periods (Supplemental Table 4). Patient characteristics by all eight unique FVC trajectory group assignments are provided in Supplemental Table 5. Differences among these trajectory groups were observed for age, calendar year period, smoking status, RF/anti-CCP seropositivity, and csDMARD use.

Prognosis in primary RA-ILD cohort by trajectory. *Mortality.* Over 5,263 patient-years of follow-up, 727 deaths occurred. Compared to the stable group, the slow progression (adjusted hazard ratio [aHR] 1.55; confidence interval [CI] 1.25, 1.91) and rapid progression groups (aHR 2.07 [CI 1.44, 2.97]) were associated with poorer survival (Figure 2; Table 2).

Respiratory hospitalization. A total of 646 respiratory hospitalizations occurred over 3,932 patient-years of follow-up. The slow progression (aHR 1.38 [CI 1.12, 1.71]) and rapid progression (aHR 1.77 [CI 1.20, 2.61]) groups were associated with a higher respiratory hospitalization risk than the stable group (Figure 2; Table 2).

Prognosis in overall RA-ILD cohort by trajectories. *Mortality.* Over 24,780 patient-years of follow-up, 3,400 deaths occurred. Trajectory group 2 was used as the reference group given its size (20.5% of patients), normal initial FVC, and stable trajectory. All four progressive groups were individually associated with poorer survival in adjusted models (Figure 3A; Table 3). The aHR (95% CI) for these groups were 1.19 (1.04, 1.37) for group 3, 1.56 (1.37, 1.78) for group 4, 2.14 (1.81, 2.53) for group 7, and 3.16 (2.45, 4.09) for group 8. None of the nonprogressive trajectory groups were associated with higher mortality risk (aHR 0.89–1.07), including group 6, which had a reduced initial FVC value. When categorizing these eight groups into a binary of progressive versus nonprogressive, the progressive group was associated with a >1.4-fold higher risk of death (aHR 1.44 [CI 1.34, 1.56]) independent of initial FVC (Figure 3B; Table 3).

Respiratory-related hospitalization. Over 19,404 patient-years of follow-up, 2,763 patients with RA-ILD experienced a respiratory-related hospitalization. Referent to trajectory group 2, all progressive FVC trajectory groups had a higher risk of respiratory-related hospitalization (Figure 3C; Table 3). The aHR (95% CI) for respiratory-related hospitalization among these groups were 1.30 (1.12, 1.52) for group 3, 1.70 (1.47, 1.96) for group 5, 2.12 (1.76, 2.55) for group 7, and 3.07 (2.32, 4.07) for group 8. None of the groups categorized as nonprogressive



Group	% of RA-ILD patients	ΔFVC % pred./year (SE)
Stable	17.5	3.17 (0.28)
Slow prog.	75.8	-2.16 (0.16)
Rapid prog.	6.7	-10.78 (0.51)

Group	% of RA-ILD patients	FVC % pred. Intercept	ΔFVC % pred./year (SE)
1	2.3	110.3	-0.19 (0.11)
2	20.5	93.5	-0.29 (0.08)
3	8.7	81.1	**-1.56 (0.14)**
4	20.1	79.1	0.21 (0.09)
5	23.9	67.3	**-1.38 (0.09)**
6	3.2	63.9	0.36 (0.09)
7	18.1	54.2	**-1.22 (0.07)**
8	3.3	40.9	**-0.73 (0.11)**

Abbreviations: FVC, forced vital capacity; RA-ILD, rheumatoid arthritis-associated interstitial lung disease; SE, standard error

Figure 1. Unique trajectories of FVC measurements among patients with RA-ILD. FVC percent predicted over time in patients with RA-ILD in the Veterans Affairs Health Care System for (A) the primary RA-ILD cohort and (B) the overall RA-ILD cohort. Clusters of patients with RA-ILD defined by unique FVC trajectories using group-based trajectory modeling. The dots represent the mean values at each time point. The percentage of patients belonging to each group, initial FVC percent predicted (overall RA-ILD cohort only), and change in FVC percent predicted over time are summarized in the tables. FVC, forced vital capacity; G, group; RA-ILD, rheumatoid arthritis–associated interstitial lung disease.

were associated with respiratory hospitalization risk. Compared to the combined nonprogressive group, those with a progressive FVC trajectory had a significantly increased risk of respiratory-related hospitalization (aHR 1.51 [CI 1.38, 1.64]; Figure 3D; Table 3).

DISCUSSION

We aimed to evaluate the heterogeneity of RA-ILD throughout its disease course by deriving unique clusters of patients with

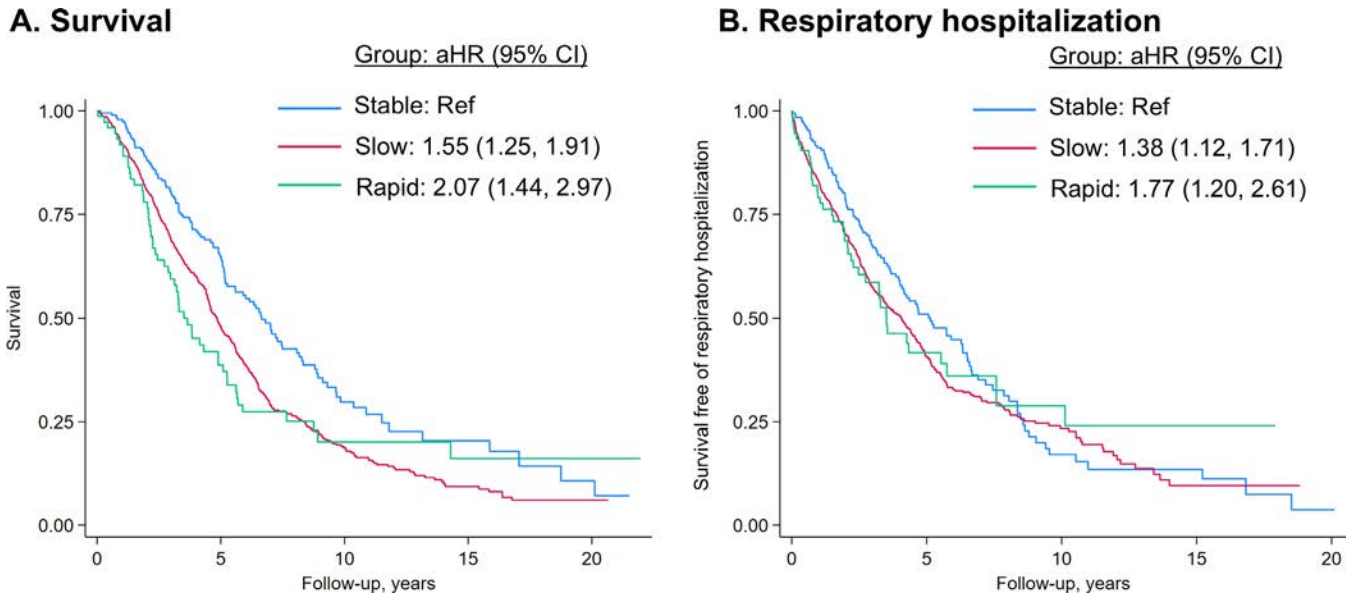


Figure 2. Kaplan–Meier curves of survival and respiratory hospitalization among patients with RA-ILD by FVC trajectory group assignment in the primary RA-ILD cohort. (A) Survival and (B) respiratory-related hospitalizations throughout the study follow-up by FVC trajectory group assignment. aHRs from multivariable models including age, sex, race, smoking status, calendar year period, comorbidity burden, and baseline FVC percent predicted. aHR, adjusted hazard ratio; CI, confidence interval; FVC, forced vital capacity; RA-ILD, rheumatoid arthritis–associated interstitial lung disease.

RA-ILD based on longitudinal FVC trajectories. Using national VA data over a >20-year period, we identified clusters of patients with RA-ILD with stable, slowly progressive, and rapidly progressive disease courses. In total, 82.5% of patients in our primary RA-ILD cohort (n = 1,092) and 54% of patients in our broader overall RA-ILD cohort (n = 5,172) had a progressive disease phenotype. Progressive FVC decline trajectory clusters were strongly associated with both death and respiratory-related hospitalization independent of initial FVC and other patient characteristics. Our findings further establish the heterogeneity of the RA-ILD disease course and emphasize the need to develop approaches to prognosticate the disease course at, or shortly after, diagnosis. With such tools, clinicians will be better equipped to make treatment and monitoring decisions in RA-ILD.

The findings from our national RA-ILD study are in line with prior single center studies that identified three to four FVC trajectory groups in RA-ILD.[12,13] One notable difference was that a higher proportion of patients with RA-ILD in our cohort had a progressive disease course (82.5% vs 44% and 68%, respectively). In our broader overall RA-ILD cohort, the number of trajectories identified was very similar to findings reported by Man et al in systemic sclerosis–associated ILD, who identified seven unique FVC trajectories.[23] In both of our RA-ILD cohorts, the rate of change was the parameter most closely associated with risk of death or hospitalization, with these progressive trajectories being associated with poor outcomes independent of initial FVC value. Findings from our overall RA-ILD cohort demonstrate that the highest risk groups were patients with RA-ILD that had the most

Table 2. Survival and respiratory hospitalization risk by FVC trajectory group assignment in primary RA-ILD cohort*

	Number of events / PY	IR (95% CI)[a]	aHR (95% CI)	*P* value
Mortality				
Overall	727 / 5,263	1.4 (1.3, 1.5)	–	–
Stable	113 / 1,113	1.0 (0.8, 1.2)	Ref.	–
Slow progression	561 / 3,814	1.5 (1.4, 1.6)	1.55 (1.25, 1.91)	<0.001
Rapid progression	53 / 337	1.6 (1.2, 2.1)	2.07 (1.44, 2.97)	<0.001
Respiratory hospitalization				
Overall	646 / 3,932	1.6 (1.5, 1.8)	–	–
Stable	117 / 832	1.4 (1.2, 1.7)	Ref.	–
Slow progression	488 / 2,842	1.7 (1.6, 1.9)	1.38 (1.12, 1.71)	0.002
Rapid progression	41 / 258	1.6 (1.2, 2.2)	1.77 (1.20, 2.61)	0.004

* Adjusted for age, sex, race, smoking status, diagnosis year period, number of baseline comorbidities, and baseline FVC percent predicted. aHR, adjusted hazard ratio; CI, confidence interval; FVC, forced vital capacity; IR, incidence rate; PY, person-years; RA-ILD, rheumatoid arthritis–associated interstitial lung disease.
[a] Per 10 PY

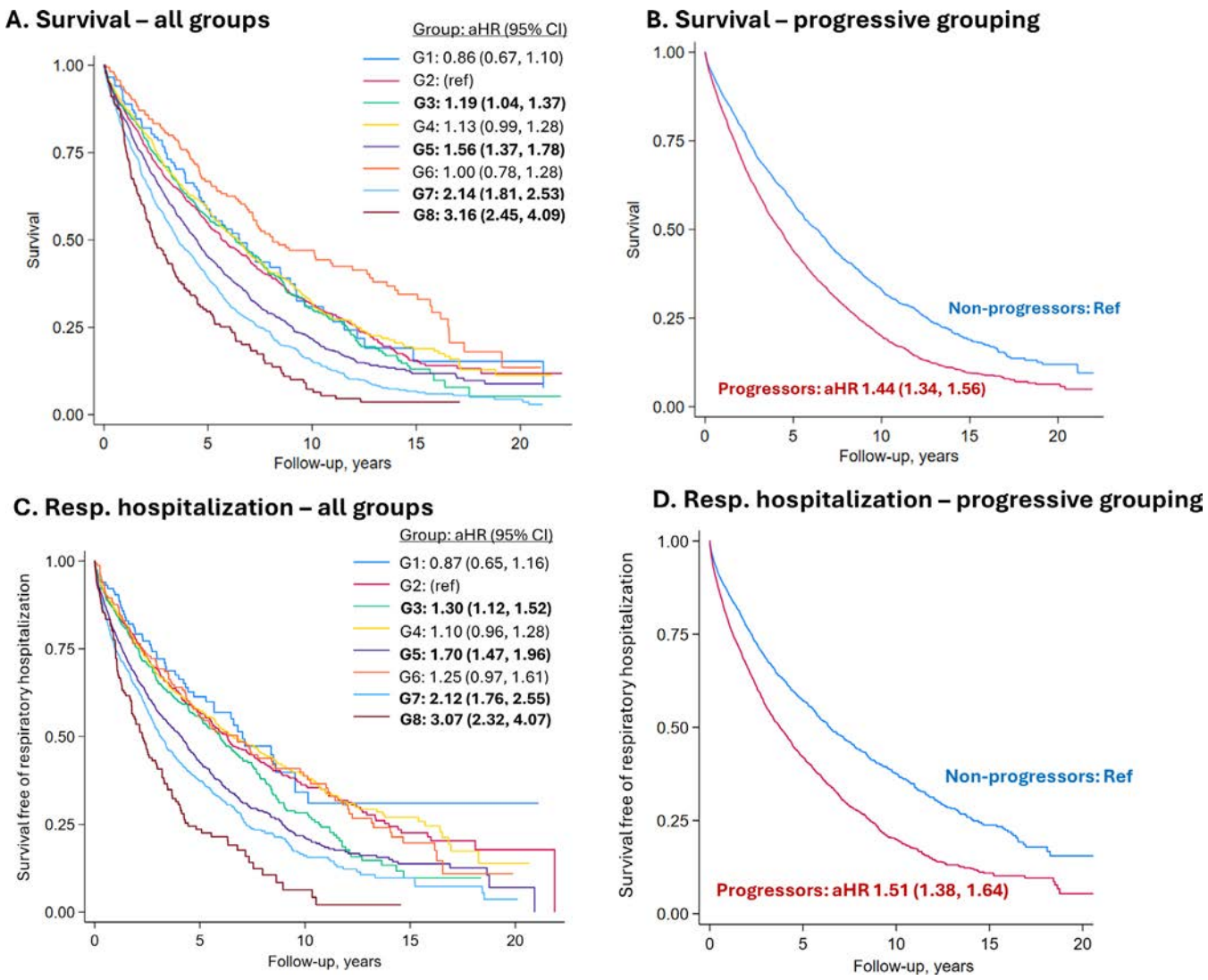


Figure 3. Kaplan–Meier curves of survival and respiratory hospitalization among patients with RA-ILD by FVC trajectory group assignment in the overall RA-ILD cohort. Survival and respiratory-related hospitalization outcomes by FVC trajectory group assignment. (A) Depicts survival in all group assignments, whereas (B) depicts survival in progressive versus nonprogressive group assignments. (C) Depicts survival free of respiratory hospitalization in all groups, whereas (D) depicts survival free of respiratory hospitalization in progressive versus nonprogressive groups. aHRs from multivariable models including age, sex, race, smoking status, calendar year period, comorbidity burden, and baseline FVC percent predicted. aHR, adjusted hazard ratio; CI, confidence interval; FVC, forced vital capacity; G, group; RA-ILD, rheumatoid arthritis–associated interstitial lung disease.

rapidly progressive decline in FVC in addition to an impaired starting FVC value (eg, groups 7 and 8). Thus, reducing patients with RA-ILD to a simple binary progressive and nonprogressive categorization may be an oversimplification of RA-ILD progression. In contrast, a single nonprogressive RA-ILD group appeared to summarize a nonprogressive disease course well, as incidence rates for death and respiratory-related hospitalization outcomes were overlapping in the four nonprogressive groups in our overall RA-ILD cohort. Our findings highlight the need for precise characterization of FVC change over time and the need for clinically feasible approaches to predict these unique disease trajectories.

There is no universally agreed upon definition for progression in RA-ILD. Clinical practice guidelines for idiopathic pulmonary fibrosis (IPF) define progressive pulmonary fibrosis based on fulfilling two of three components: worsening respiratory symptoms, pulmonary physiology, and radiologic findings.[24] The FVC component of pulmonary physiology is defined based on an absolute decrease of ≥5% predicted over a 1-year follow-up period. Clinical trials have used alternative definitions as part of their eligibility criteria and study outcomes. The INBUILD trial evaluating nintedanib in progressive pulmonary fibrosis required evidence of progression for study eligibility, using a definition of relative decline in FVC precent predicted by ≥10% or ≥5% plus worsening

Table 3. Survival and respiratory hospitalization risk by FVC trajectory group in overall RA-ILD cohort*

	Number of events / PY	IR (95% CI)[a]	aHR (95% CI)	*P* value
Mortality				
Overall	3,400 / 24,780	1.4 (1.3, 1.4)	–	–
All groups				
Group 1	72 / 664	1.1 (0.9, 1.4)	0.86 (0.67, 1.10)	0.23
Group 2	614 / 5,094	1.2 (1.1, 1.3)	Ref.	–
Group 3[b]	299 / 2,488	1.2 (1.1, 1.3)	1.19 (1.04, 1.37)	0.01
Group 4	605 / 5,419	1.1 (1.0, 1.2)	1.13 (0.99, 1.28)	0.07
Group 5[b]	841 / 5,463	1.5 (1.4, 1.6)	1.56 (1.37, 1.78)	<0.001
Group 6	98 / 1,233	0.8 (0.7, 1.0)	1.00 (0.78, 1.28)	0.97
Group 7[b]	727 / 3,841	1.9 (1.8, 2.0)	2.14 (1.81, 2.53)	<0.001
Group 8[b]	144 / 577	2.5 (2.1, 2.9)	3.16 (2.45, 4.09)	<0.001
Progressive grouping				
Nonprogressive	1,389 / 12,410	1.1 (1.1, 1.2)	Ref.	–
Progressive	2,011 / 12,369	1.6 (1.6, 1.7)	1.44 (1.34, 1.56)	<0.001
Respiratory hospitalization				
Overall	2,763 / 19,404	1.4 (1.4, 1.5)	–	–
All groups				
Group 1	53 / 539	1.0 (0.8, 1.3)	0.87 (0.65, 1.16)	0.34
Group 2	473 / 4,204	1.1 (1.0, 1.2)	Ref.	–
Group 3[b]	252 / 1,912	1.3 (1.2, 1.5)	1.30 (1.12, 1.52)	0.001
Group 4	480 / 4,421	1.1 (1.0, 1.2)	1.10 (0.96, 1.28)	0.17
Group 5[b]	717 / 4,210	1.7 (1.6, 1.8)	1.70 (1.47, 1.96)	<0.001
Group 6	101 / 926	1.1 (0.9, 1.3)	1.25 (0.97, 1.61)	0.09
Group 7[b]	568 / 2,782	2.0 (1.9, 2.2)	2.12 (1.76, 2.55)	<0.001
Group 8[b]	119 / 410	2.9 (2.4, 3.5)	3.07 (2.32, 4.07)	<0.001
Progressive grouping				
Nonprogressive	1,107 / 10,090	1.1 (1.0, 1.2)	Ref.	–
Progressive	1,656 / 9,314	1.8 (1.7, 1.9)	1.51 (1.38, 1.64)	<0.001

* Adjusted for age, sex, race, smoking status, diagnosis year period, number of baseline comorbidities, and baseline FVC percent predicted. aHR, adjusted hazard ratio; CI, confidence interval; FVC, forced vital capacity; IR, incidence rate; PY, person-years; RA-ILD, rheumatoid arthritis–associated interstitial lung disease.
[a] IR per 10 PY
[b] Indicates a group categorized as progressive

symptoms or imaging findings over the prior 24 months.[11] The TRAIL1 trial evaluating pirfenidone in RA-ILD selected a primary study outcome of progression based on decline in FVC percent by ≥10% or death.[25] Compared to these prior definitions of progression, most of our progressive RA-ILD groups had FVC change rates that were well below these thresholds, yet these groups still experienced a higher risk of death and respiratory-related hospitalization during follow-up. This suggests that reliance on the higher progression thresholds used as part of prior regulatory trials may be overly stringent to inform real-world treatment decisions as patients with slow progression may experience improved disease-related outcomes with treatment modifications.

Heterogeneity in RA-ILD poses a major challenge to management. This study identified unique disease trajectories based on FVC values alone, but FVC variability represents only one contributor to RA-ILD heterogeneity alongside ILD pattern and genetic predisposition, among others. Recognizing the substantial heterogeneity in RA-ILD, precise disease subsetting has the potential to guide management decisions. However, what factors are most useful for separating these subsets is not known. In this study, we found standard demographic and clinical factors were only modestly associated with different FVC trajectories. This suggests variability in FVC trajectory is more likely related to heterogeneity in other disease-related factors and highlights the need for more advanced phenotyping, such as broader autoantibody profiling or proteomic and radiomic evaluation, of RA-ILD to identify factors underlying these unique disease trajectories. Supporting this, translational studies have identified unique peripheral blood biomarker and proteomic signatures associated with the presence of RA-ILD.[26,27] A small proof of concept study evaluated whole lung radiomic analysis in 30 patients with RA-ILD and found that a subset of the 22 unique radiomic features extracted predicted overall survival.[28] Further, Humphries and colleagues used data-driven texture analysis for chest CT quantification of fibrosis in two independent RA-ILD cohorts to predict RA-ILD survival.[29] The discovery and validation of such approaches and their application to clinical care early in the RA-ILD course could prove to be highly informative for forecasting disease trajectories and fostering earlier intervention, rather than the current status quo of waiting for disease progression to occur.

Limitations to the study include the potential lack of generalizability given the male predominance of the cohort. The high frequency of comorbid COPD likely reflects the high cigarette

smoking rates and may also impact generalizability. Our cohort consisted of both prevalent and incident RA-ILD cases, which may have deemphasized FVC trajectories during the early disease period. FVC values were obtained through real-world clinical care, which led to variable testing frequency and testing intervals. Although we used validated algorithms and NLP tools,[18,19] there may be misclassification of RA-ILD status and FVC values. ILD pattern was not available in our data. Future studies are needed to understand how RA-ILD FVC trajectories may differ based on ILD pattern. Our approach for trajectory modeling grouped patients with RA-ILD by the entirety of their FVC data into single trajectory over time. With the current study design and time period evaluated, we were not positioned to evaluate how treatment modifications (eg, DMARDs and/or antifibrotics), RA disease flares, ILD flares, or environmental exposures occurring during the disease course may have influenced FVC trajectory. FVC values and the clinical events of death and respiratory hospitalization were assessed over the same follow-up period, which could result in immortal time bias, though our findings of FVC decline associating with mortality and respiratory hospitalization risk are consistent with other studies.[30]

In conclusion, we identified unique clusters of patients with RA-ILD based on FVC trajectories using large, real-world data over a >20-year period. These clusters were defined by variability in the rate of decline throughout the follow-up period as well as initial FVC values. FVC trajectory clusters that reflected progression were strongly associated with both survival and respiratory-related hospitalization risk, despite most progressive groups not reaching thresholds for progression used in IPF clinical practice guidelines or clinical trials. These findings illustrate the heterogeneity of the RA-ILD disease course and highlight the poor prognosis even for those who experience "slow" ILD progression. Future studies are needed to develop methods that can better distinguish unique disease subsets early in the disease course to inform proactive management decisions.

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 England 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

Boehringer-Ingelheim had no role in the study design or in the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Boehringer-Ingelheim.

REFERENCES

1. Gabbay E, Tarala R, Will R, et al. Interstitial lung disease in recent onset rheumatoid arthritis. Am J Respir Crit Care Med 1997;156(2 Pt 1):528–535.
2. Natalini JG, Swigris JJ, Morisset J, et al. Understanding the determinants of health-related quality of life in rheumatoid arthritis-associated interstitial lung disease. Respir Med 2017;127:1–6.
3. Brooks R, Baker JF, Yang Y, et al. The impact of disease severity measures on survival in U.S. veterans with rheumatoid arthritis-associated interstitial lung disease. Rheumatology (Oxford) 2022;61(12):4667–4677.
4. Raimundo K, Solomon JJ, Olson AL, et al. Rheumatoid arthritis-interstitial lung disease in the United States: prevalence, incidence, and healthcare costs and mortality. J Rheumatol 2019;46(4):360–369.
5. Johnson SR, Bernstein EJ, Bolster MB, et al. 2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) Guideline for the Treatment of Interstitial Lung Disease in People with Systemic Autoimmune Rheumatic Diseases. Arthritis Rheumatol 2024;76(8):1182–1200.
6. Joy GM, Arbiv OA, Wong CK, et al. Prevalence, imaging patterns and risk factors of interstitial lung disease in connective tissue disease: a systematic review and meta-analysis. Eur Respir Rev 2023;32(167):220210.
7. Juge PA, Lee JS, Ebstein E, et al. MUC5B promoter variant and rheumatoid arthritis with interstitial lung disease. N Engl J Med 2018;379(23):2209–2219.
8. Wheeler AM, Baker JF, Poole JA, et al. Genetic, social, and environmental risk factors in rheumatoid arthritis-associated interstitial lung disease. Semin Arthritis Rheum 2022;57:152098.
9. Singh N, Varghese J, England BR, et al. Impact of the pattern of interstitial lung disease on mortality in rheumatoid arthritis: a systematic literature review and meta-analysis. Semin Arthritis Rheum 2019;49(3):358–365.
10. Matson SM, Baqir M, Moua T, et al. Treatment outcomes for rheumatoid arthritis-associated interstitial lung disease: a real-world, multisite study of the impact of immunosuppression on pulmonary function trajectory. Chest 2023;163(4):861–869.
11. Flaherty KR, Wells AU, Cottin V, et al; INBUILD Trial Investigators. Nintedanib in progressive fibrosing interstitial lung diseases. N Engl J Med 2019;381(18):1718–1727.
12. Chang SH, Lee JS, Ha YJ, et al. Lung function trajectory of rheumatoid arthritis-associated interstitial lung disease. Rheumatology (Oxford) 2023;62(9):3014–3024.
13. Venkat RK, Hayashi K, Juge PA, et al. Forced vital capacity trajectories and risk of lung transplant and ILD-related mortality among patients with rheumatoid arthritis-associated interstitial lung disease. Clin Rheumatol 2024;43(8):2453–2466.
14. Johnson TM, Mahabir CA, Yang Y, et al. Aortic stenosis risk in rheumatoid arthritis. JAMA Intern Med 2023;183(9):973–981.
15. Johnson TM, Yang Y, Roul P, et al. A narrowing mortality gap: temporal trends of cause-specific mortality in a national matched cohort study in US veterans with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2023;75(8):1648–1658.
16. Brooks RT, Luedders B, Wheeler A, et al. The risk of lung cancer in rheumatoid arthritis and rheumatoid arthritis-associated interstitial lung disease. Arthritis Rheumatol 2024;76(12):1730–1738.
17. Chung CP, Rohan P, Krishnaswami S, et al. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. Vaccine 2013;31(suppl 10):K41–K61.

18. England BR, Roul P, Mahajan TD, et al. Performance of administrative algorithms to identify interstitial lung disease in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2020;72(10):1392–1403.
19. England BR, Roul P, Yang Y, et al. Extracting forced vital capacity from the electronic health record through natural language processing in rheumatoid arthritis-associated interstitial lung disease. Pharmacoepidemiol Drug Saf 2024;33(1):e5744.
20. Sohn M-W, Arnold N, Maynard C, et al. Accuracy and completeness of mortality data in the Department of Veterans Affairs. Popul Health Metr 2006;4(1):2.
21. England BR, Roul P, Yang Y, et al. Burden and trajectory of multimorbidity in rheumatoid arthritis: a matched cohort study from 2006 to 2015. Ann Rheum Dis 2021;80(3):286–292.
22. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 2010;6(1):109–138.
23. Man A, Davidyock T, Ferguson LT, et al. Changes in forced vital capacity over time in systemic sclerosis: application of group-based trajectory modelling. Rheumatology (Oxford) 2015;54(8):1464–1471.
24. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med 2022;205(9):e18–e47.
25. Solomon JJ, Danoff SK, Woodhead FA, et al; TRAIL1 Network Investigators. Safety, tolerability, and efficacy of pirfenidone in patients with rheumatoid arthritis-associated interstitial lung disease: a randomised, double-blind, placebo-controlled, phase 2 study. Lancet Respir Med 2023;11(1):87–96.
26. Wheeler AM, Baker JF, Riley TR IV, et al. Peripheral biomarker signatures in rheumatoid arthritis-associated interstitial lung disease. Arthritis Rheumatol 2025;77(8):971–980.
27. Wu X, Jeong Y, Poli de Frías S, et al. Serum proteomic profiling of rheumatoid arthritis-interstitial lung disease with a comparison to idiopathic pulmonary fibrosis. Thorax 2022;77(10):1041–1044.
28. Venerito V, Manfredi A, Lopalco G, et al. Radiomics to predict the mortality of patients with rheumatoid arthritis-associated interstitial lung disease: a proof-of-concept study. Front Med (Lausanne) 2023;9:1069486.
29. Humphries SM, Adegunsoye A, Demoruelle MK, et al. Quantitative CT scan analysis in rheumatoid arthritis-related interstitial lung disease. Chest 2025;167(5):1428–1439.
30. Paterniti MO, Bi Y, Rekić D, et al. Acute exacerbation and decline in forced vital capacity are associated with increased mortality in idiopathic pulmonary fibrosis. Ann Am Thorac Soc 2017;14(9):1395–1402.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 259–269
DOI 10.1002/acr.25618

American College of Rheumatology
Empowering Rheumatology Professionals

Improvements in Health-Related Quality of Life With Treat-to-Target Urate-Lowering Therapy in Gout: A Post Hoc Analysis of a Randomized Multicenter Trial

Austin Barry,[1] Bryant R. England,[1] Harlan Sayles,[2] Lindsay N. Helget,[1] Maria Androsenko,[3] Hongsheng Wu,[4] Kaleb Michaud,[5] Bridget Kramer,[1] Jeff A. Newcomb,[1] Mary T. Brophy,[6] Anne Davis-Karim,[7] Ryan Ferguson,[3] Michael H. Pillinger,[8] Tuhina Neogi,[9] Paul M. Palevsky,[10] James R. O'Dell,[1] and Ted R. Mikuls[1]

Objective. Although treat-to-target urate-lowering therapy (ULT) is endorsed as best practice in gout management, limited data exist on its impact on health-related quality of life (HRQoL). We assessed the impact of treat-to-target ULT on HRQoL among participants receiving protocolized gout care, identifying factors associated with HRQoL and HRQoL change.

Methods. This was a post hoc analysis of a 72-week randomized trial, pooling data from allopurinol and febuxostat treatment arms. The Veterans RAND-12 Item Health Survey and EuroQol 5-Dimension 3-Level (EQ-5D-3L) were administered at baseline and at 24, 48, and 72 weeks. HRQoL changes over follow-up were examined using paired *t*-tests. Factors associated with baseline HRQoL were evaluated using multivariable linear regression. General estimating equations were used to identify determinants of HRQoL change over follow-up.

Results. Participants (N = 878) in this analysis were 98.9% male, had a mean age of 62.4 years, and 67.4% self-reported White race. HRQoL scores overall, and particularly the domains of physical function, mobility and pain, improved significantly over 72 weeks ($P < 0.001$) with improvements noted by 24 weeks. Poorer enrollment HRQoL was associated with younger age, non-White race, tophi (for EQ-5D-3L), higher serum urate level, and greater comorbidity. Baseline factors associated with HRQoL improvements over 72 weeks of ULT included lower C-reactive protein level and lower comorbidity scores with similar changes observed by ULT assignment.

Conclusion. Treat-to-target ULT in gout is accompanied by HRQoL improvements evident by 24 weeks and sustained through 72 weeks. HRQoL gains with treat-to-target ULT are most prominent in the domains of physical function, mobility, and pain and are greatest in those with lower baseline levels of inflammation and comorbidity.

INTRODUCTION

Gout is characterized by acute episodes of arthritis (or flares) separated by intercritical periods and is the most common form of inflammatory arthritis, affecting up to 5% of the population and nearly 12 million people in the United States alone.[1] Consequently, gout is associated with significant reductions in health-related quality of life (HRQoL) and work participation.[2–6] Recognizing the detrimental impact that gout exerts on HRQoL, the Outcome Measures in Rheumatology consortium

ClinicalTrials.gov identifier: NCT02579096.

Supported by the Veterans Affairs (VA) Cooperative Studies Program of the Department of Veterans Affairs Office of Research & Development. Dr England's work was supported by a VA Clinical Science Research and Development grant IK2 (CX002203) and the Rheumatology Research Foundation. Dr Michaud's work was supported by the Rheumatology Research Foundation. Dr Pillinger's work was supported by the National Institutes of Health (NIH; grant 1UL1-TR-001445) and the VA (I01 CX002358). Dr Neogi's work was supported by the NIH (grants K24-AR-070892, P30-AR-072571). Dr Mikuls' work was supported by the VA (Merit grant BX003635), US Department of Defense (grant PR200793), and the NIH (grant U54-GM-115458).

[1]Austin Barry, MD, Bryant R. England, MD, PhD, Lindsay N. Helget, MD, Bridget Kramer, RN, BSN, Jeff A. Newcomb, BS, James R. O'Dell, MD, Ted R. Mikuls, MD, MSPH: VA Nebraska-Western Iowa Health Care System and University of Nebraska Medical Center, Omaha; [2]Harlan Sayles, PhD: University of Nebraska Medical Center, Omaha; [3]Maria Androsenko, MS, Ryan Ferguson, ScD: VA Boston Cooperative Studies Program Coordinating Center, Boston, Massachusetts; [4]Hongsheng Wu, PhD: VA Boston Cooperative Studies Program Coordinating Center and Wentworth Institute of Technology, Boston, Massachusetts; [5]Kaleb Michaud, PhD: VA Nebraska-Western Iowa Health Care System and University of Nebraska Medical Center, Omaha, and Forward Databank, Wichita, Kansas; [6]Mary T. Brophy, MD: VA Boston Cooperative Studies Program Coordinating Center and VA Boston Health Care System, Boston University, Boston, Massachusetts; [7]Anne Davis-Karim, PharmD: VA Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque, New Mexico; [8]Michael H. Pillinger, MD: VA New York Harbor Health Care System and NYU Grossman School of Medicine, New York, New York; [9]Tuhina Neogi, MD, PhD: Boston University Chobanian &

SIGNIFICANCE & INNOVATIONS

- Despite being recommended in clinical practice guidelines, the impact of treat-to-target urate-lowering therapy on health-related quality of life (HRQoL) in gout is poorly understood.
- Younger age, non-White race, tophi, higher serum urate, and comorbidity were associated with poorer HRQoL in trial participants with gout at enrollment.
- Treat-to-target urate-lowering therapy led to significant improvements in HRQoL over 72 weeks, gains that were evident as early as 24 weeks.
- The physical component score of the Veteran RAND 12-item and EuroQol 5-Dimension 3-Level domains of pain and mobility demonstrated significant improvements with treat-to-target urate-lowering therapy.
- Participants with gout accompanied by fewer baseline comorbidities and lower C-reactive protein levels were most likely to demonstrate HRQoL improvements with treat-to-target urate-lowering therapy.

has endorsed the routine assessment of HRQoL as a patient-reported outcome in trials investigating the efficacy of urate-lowering therapy in gout.[7]

Acknowledging the central role of hyperuricemia in gout pathogenesis, urate-lowering therapy is a cornerstone in gout treatment and demonstrates beneficial effects that include the lowering of serum urate concentration, promoting the dissolution of monosodium urate crystals, reducing or eliminating gout flares, resolving tophi, and preventing joint damage. International subspecialty societies, including both the American College of Rheumatology (ACR) and EULAR, recommend a treat-to-target approach to urate-lowering therapy with a goal of achieving and maintaining a serum urate of <6 mg/dL.[8,9] To date, most studies linking gout with reductions in HRQoL have been cross-sectional in design.[2] Thus, despite the well-documented associations of gout with reduced HRQoL and published recommendations strongly endorsing treat-to-target urate-lowering therapy, there remain only limited data exploring how this best-practice management approach influences HRQoL outcomes.[10–12]

To further assess the HRQoL changes that accompany treat-to-target urate-lowering therapy, we conducted a post hoc analysis of the STOP Gout trial (ClinicalTrials.gov identifier NCT02579096), a comparative effectiveness study that examined the efficacy of allopurinol and febuxostat with both agents administered following a treat-to-target protocol.[13] In this analysis, data from the allopurinol and febuxostat treatment arms were pooled to test the hypothesis that study participants with gout receiving treat-to-target urate-lowering therapy would yield significant improvements in HRQoL over follow-up as assessed by the Veterans RAND 12-item (VR-12) and the EuroQol 5-Dimension 3-Level (EQ-5D-3L) questionnaires. We also sought to identify participant characteristics, including urate-lowering therapy treatment assignment, that might predict such improvements, by testing the additional hypotheses that the improvement in HRQoL measures would be associated with previously documented reductions in gout flares and the achievement of serum urate goal.

PATIENTS AND METHODS

Data source. The STOP Gout study was a multicenter, randomized, double-blind, placebo-controlled, 72-week noninferiority trial comparing allopurinol with febuxostat in 940 participants with a primary outcome of flare occurring during the final study phase (weeks 49–72).[13] The trial was conducted at 19 Veterans Affairs (VA) sites and two academic medical centers. By design, approximately one-third of participants had Stage 3 chronic kidney disease (CKD). Individuals with more advanced CKD (stages 4 or 5) were excluded. Participants with prior allopurinol use (≤300 mg/day) but not at serum urate goal were also included in the trial. The treatment protocol consisted of three phases of equal duration: urate-lowering therapy initiation and titration (phase 1: weeks 0–24), maintenance (phase 2: weeks 25–48), and observation (phase 3: weeks 49–72) (Figure 1).

Survey data were available for the EQ-5D-3L Index and VR-12 in 877 and 875 participants, respectively, at baseline; 748 and 759 at 24 weeks; 761 and 750 at 48 weeks; and 691 and 692 at 72 weeks. Participants with gout and serum urate level >6.8 mg/dL were randomly assigned to allopurinol or febuxostat. For those not taking allopurinol at enrollment, agents were initiated at daily doses of 100 mg for allopurinol and 40 mg for febuxostat with doses gradually escalated to achieve a serum urate level <6.0 mg/dL (<5.0mg/dL if tophi were present) or until the maximal allowable urate-lowering therapy dose was reached. Participants taking allopurinol ≤300 mg daily before the trial were eligible and continued taking their pretrial (blinded) allopurinol dose if relevant and randomly assigned to that arm, with dose titration delayed for those receiving 200 mg (starting at week 6) or 300 mg (starting at week 9). The dose titration schedule[13] during phase 1 followed the 2012 ACR gout management guidelines,[14,15] with maximum daily doses of 800 mg of

Avedisian School of Medicine, Boston, Massachusetts; [10]Paul M. Palevsky, MD: VA Pittsburgh Health Care System and University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.

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.25618).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25618.

Address correspondence via email to Ted R. Mikuls, MD, MSPH, at tmikuls@unmc.edu.

Submitted for publication February 6, 2025; accepted in revised form July 22, 2025.

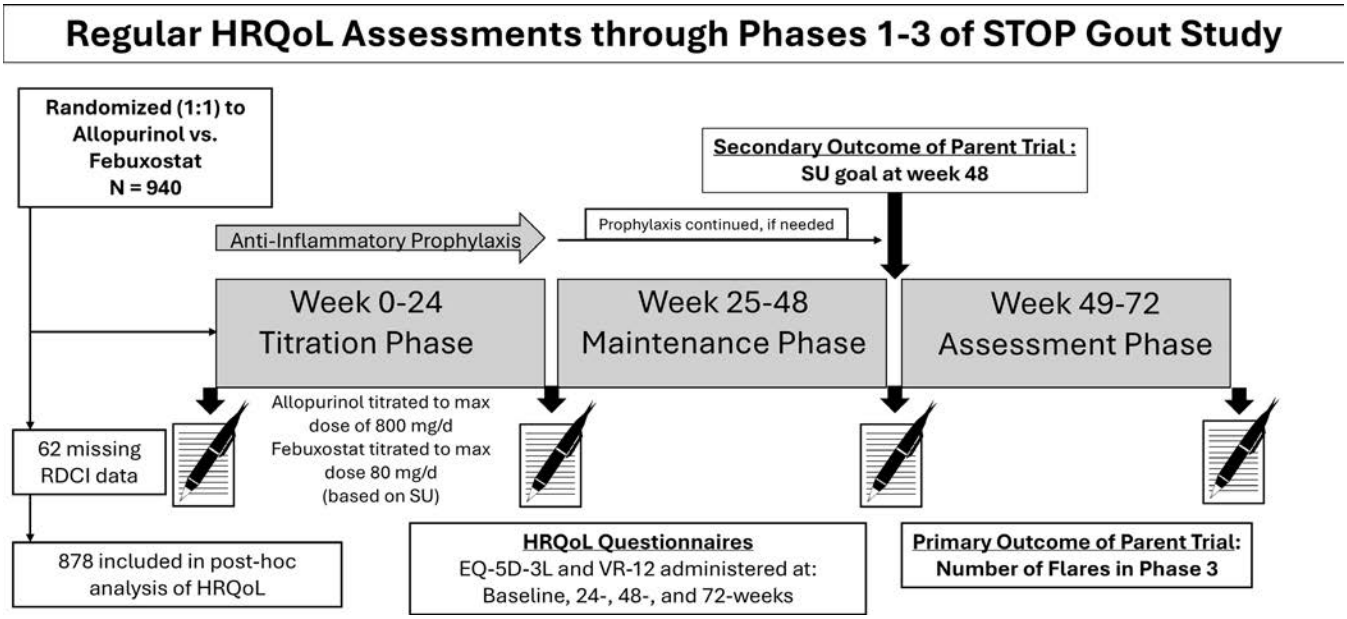


Figure 1. General structure and endpoints of the STOP Gout study with routine HRQoL assessments. During phase 1 (weeks 0–24), urate-lowering therapy was titrated according to 2012 American College of Rheumatology guidelines. In phase 2 (weeks 25–48), the urate-lowering therapy dose was adjusted as needed. In phase 3 (weeks 49–72), the urate-lowering therapy dose was stable, and participants were monitored for flare. EQ-5D-3L, EuroQol 5-Dimension 3-Level; HRQoL, health-related quality of life; RDCI, Rheumatic Disease Comorbidity Index; SU, serum urate; VR-12, Veterans RAND 12-item.

allopurinol and 120 mg of febuxostat. The maximum allowable daily febuxostat dose was reduced from 120 mg to 80 mg during the study, in February 2019, at the request of the US Food and Drug Administration following a modification of febuxostat labeling to include a black box warning detailing cardiovascular safety concerns.[16] All participants received guideline-directed anti-inflammatory prophylaxis per 2012 ACR guidelines[15] with colchicine, naproxen 250 mg twice daily, or prednisone ≤10 mg daily per investigator discretion that extended through the entirety of phase 1. Additional prophylaxis was allowed during phase 2 with discontinuation mandatory before the end of this phase (by the end of week 48).

Adherence to urate-lowering therapy was assessed throughout the trial using participant-maintained diaries. Adherence data were available for 83% of participants, with compliance (defined as taking the assigned study urate-lowering therapy ≥80% of days) reported in 84.2%.[17] The trial indicated that allopurinol was noninferior to febuxostat in controlling flares, with similar achievement of serum urate goals (80% overall achieved serum urate goal) and no significant differences in serious adverse events.[13]

HRQoL questionnaires. Two questionnaires, the VR-12 and EQ-5D-3L, were administered to STOP Gout participants at baseline and at 24, 48, and 72 weeks. The VR-12 was developed and modified from the original RAND version of the 12-Item Health Survey version 1.0 (also known as the 12-item Short Form [SF-12]).[18–23] Similar to the SF-12, VR-12 results are reported in two scores: a physical component score (PCS) and mental component score (MCS), each standardized to the VA population mean ± SD of 50 ± 10 with higher scores reflecting better HRQoL. Derived from the VR-12, the Veterans RAND 6-Dimension (VR-6D) survey is a single HRQoL utility index with a range of 0 to 1 with 0 representing death and 1 perfect health.[24] Similarly, the EQ-5D-3L includes five items, each with three possible responses, addressing the HRQoL domains of mobility, self-care, activity, pain, and mental health. The EQ-5D-3L with US-based tariffs produces an overall index score with range 0 to 1 with 0 representing death and 1 representing perfect health. In contrast, higher scores in individual EQ-5D-3L domains (range 1–3) reflect worse HRQoL. To facilitate data interpretation, individual EQ-5D-3L domain scores were reversed such that higher scores reported herein denote better HRQoL.

Analyses. Changes in VR-6D and EQ-5D-3L index scores from baseline to 72 weeks among all participants (allopurinol and febuxostat arms pooled) served as the co-primary endpoints for this post hoc analysis. Descriptive statistics were used to compare participant characteristics by urate-lowering therapy treatment assignment (ie, allopurinol vs febuxostat) as previously reported[13] in addition to comparing characteristics of those completing the 72-week study and those dropping out at earlier time points. In initial cross-sectional analyses, factors associated with baseline HRQoL in participants with gout were assessed using

univariable linear regression. All variables with $P < 0.1$ in univariable analyses were then included in a multivariable linear regression model. In these analyses, beta coefficients (β) and 95% confidence intervals (CIs) were generated and multiplied by a factor of 100 to facilitate interpretation. Variables examined included demographics (age, sex, race, ethnicity, and urban/rural residence), urate-lowering therapy assignment, gout duration, presence of tophi, prior allopurinol receipt, high-sensitivity C-reactive protein (hs-CRP) concentration (mg/L), serum urate (mg/dL), presence of Stage 3 CKD, overall comorbidity burden as assessed by the Rheumatic Disease Comorbidity Index (RDCI),[25] diuretic use, body mass index (kg/m^2), and regular alcohol use. The presence of component conditions making up the RDCI were identified using diagnostic codes from linked electronic health record data contained within the VA Corporate Data Warehouse. Both race and ethnicity were self-reported by participants in response to open-ended questions posed by study personnel. Non-VA study participants with missing RDCI values (n = 62) were excluded from this analysis in addition to the longitudinal analyses detailed below, leaving 878 total participants.

Changes in the VR-6D survey, EQ-5D-3L index, and component scores between baseline and 72-week values were examined using paired *t*-tests. Radar plots[26] were used to graphically demonstrate changes in individual EQ-5D-3L components. Participants with missing HRQoL values from either of the two relevant time points were excluded from analyses (Figure 1). Among those with baseline EQ-5D-3L data (n = 877), 14.7% (n = 129) were missing data from 24 weeks, 13.2% (n = 116) from 48 weeks, and 21.2% (n = 186) from 72 weeks. For those with baseline VR-12 data (n = 875), 13.3% were missing data from 24 and 48 weeks (n = 116 for both) and 20.9% (n = 183) were missing data from 72 weeks. To further explore whether participant attrition impacted changes in HRQoL observed over follow-up, we performed an additional analysis restricted to individuals with complete HRQoL data available at all study time points. A standardized mean gain (SMG) was generated to quantify the magnitude of change observed for each measure over follow-up by calculating the mean change between baseline and a later period divided by the SD at baseline. Values of 0.20 were indicative of a small effect, 0.50 a medium effect, and >0.80 a large effect.[27] General estimating equations (GEEs) were used to examine associations of participant factors with change in HRQoL from baseline to 24, 48, and 72 weeks. In addition to the same covariates as detailed above in cross-sectional analyses, additional variables examined in GEE models included baseline HRQoL values, flare counts during the preceding study phase, the achievement of serum urate goal in the previous phase as determined by blood draws taken at the completion of the phase, and anti-inflammatory prophylaxis use (colchicine vs other).

Ethical considerations. The STOP Gout study was approved by the VA Central Institutional Review Board, and all participants provided written informed consent before enrollment. The STOP Gout Steering Committee provided approval for the conduct of this post hoc analysis.

RESULTS

Characteristics of the 878 study participants included in the analysis are summarized in Table 1. The cohort was primarily male (98.9%), had self-reported White race (67.4%), had a mean age of 62.4 years, had a mean serum urate of 8.53 mg/dL, and had a mean disease duration of 10 years at enrollment. Approximately one-third (37.8%) of participants had CKD Stage 3, and 16.6% had tophi. Approximately one-third (37%) of participants had previously used allopurinol. Baseline HRQoL measures are shown in Table 2. Overall, participants completing the study were similar to those dropping out at earlier time points except that noncompleters had a worse baseline EQ-5D-3L score (0.66 vs 0.71) and were more likely to live in an urban area (Supplementary Table S1).

Univariable associations of participant characteristics with baseline/enrollment HRQoL are summarized in Supplementary Table S2. In multivariable models, factors independently associated with worse enrollment/baseline HRQoL included younger age, Black/African American race (vs White), presence

Table 1. Baseline characteristics of study participants (N = 878)*

Baseline characteristics	Data
Demographics	
Age, mean (SD), y	62.4 (12.2)
Male, %	98.9
Race, %	
White/Caucasian	67.4
Black/African American	22.2
Other	10.4
Hispanic ethnicity, %	5.0
Urban residence, %	73.3
Gout-related factors	
Serum urate level, mean (SD), mg/dL	8.53 (1.37)
C-reactive protein, mean (SD), mg/L	8.98 (17.11)
Gout duration, mean (SD), y	10.0 (11.1)
Prior allopurinol use, %	37.0
Presence of tophi, %	16.6
Treatment assigned, %	
Allopurinol	50.0
Febuxostat	50.0
Comorbidities	
Chronic kidney disease, Stage 3, %	37.8
Diuretic use, %	43.5
RDCI	2.2 (1.57)
BMI (kg/m^2), %	
<25 (healthy)	5.3
25 ≤ BMI < 30 (overweight)	27.6
30 ≤ BMI < 35 (obese)	31.7
≥35 (morbidly obese)	35.4
Alcohol use, %	54.5

* "Other" race combines Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and "none of the above."
BMI, body mass index; RDCI, Rheumatic Disease Comorbidity Index.

Table 2. VR-12 (derived VR-6D) and EQ-5D-3L scores in participants receiving treat-to-target urate-lowering therapy at baseline, 24, 48, and 72 weeks*

	Initial	24 weeks	48 weeks	72 weeks
Overall HRQoL measures				
VR-6D, mean (SD)	0.66 (0.12)	0.70 (0.12)	0.70 (0.12)	0.70 (0.12)
Change from initial, mean (SD)	–	0.04 (0.10)	0.03 (0.10)	0.04 (0.10)
P value	–	<0.001	<0.001	<0.001
Effect size, SMG	–	0.31	0.28	0.30
n	875	759	750	692
EQ-5D-3L index, mean (SD)	0.70 (0.22)	0.75 (0.21)	0.75 (0.22)	0.74 (0.22)
Change from initial, mean (SD)	–	0.05 (0.18)	0.04 (0.19)	0.03 (0.19)
P value	–	<0.001	<0.001	<0.001
Effect size, SMG	–	0.22	0.18	0.12
n	877	748	761	692
VR-12 component scores				
VR-12 Physical (PCS), mean (SD)	36.8 (10.9)	40.7 (11.5)	40.6 (11.5)	40.9 (11.5)
Change from initial, mean (SD)	–	3.9 (9.8)	3.6 (9.9)	3.7 (10.4)
P value	–	<0.001	<0.001	<0.001
Effect size, SMG	–	0.35	0.32	0.33
VR-12 Mental (MCS), mean (SD)	51.3 (11.5)	51.9 (11.0)	52.1 (10.7)	52.0 (10.8)
Change from initial, mean (SD)	–	0.4 (9.7)	0.4 (10.3)	0.4 (10.7)
P value	–	0.28	0.35	0.38
Effect size, SMG	–	0.03	0.03	0.03
EQ-5D-3L dimension scores				
Mobility, mean (SD)	2.45 (0.50)	2.58 (0.50)	2.57 (0.50)	2.54 (0.51)
Change from initial, mean (SD)	–	0.13 (0.53)	0.12 (0.53)	0.09 (0.52)
P value	–	<0.001	<0.001	<0.001
Effect size, SMG	–	0.26	0.24	0.18
Self-care, mean (SD)	2.87 (0.34)	2.88 (0.34)	2.87 (0.35)	2.86 (0.38)
Change from initial, mean (SD)	–	0.00 (0.35)	−0.01 (0.34)	−0.02 (0.36)
P value	–	0.75	0.52	0.10
Effect size, SMG	–	0.01	−0.02	−0.07
Daily activities, mean (SD)	2.55 (0.55)	2.64 (0.51)	2.61 (0.54)	2.59 (0.53)
Change from initial, mean (SD)	–	0.08 (0.55)	0.06 (0.59)	0.04 (0.59)
P value	–	<0.001	0.009	0.30
Effect size, SMG	–	0.15	0.10	0.04
Pain, mean (SD)	2.19 (0.58)	2.39 (0.58)	2.37 (0.58)	2.36 (0.58)
Change from initial, mean (SD)	–	0.18 (0.64)	0.16 (0.63)	0.14 (0.62)
P value	–	<0.001	<0.001	<0.001
Effect size, SMG	–	0.32	0.28	0.25
Mood, mean (SD)	2.67 (0.53)	2.68 (0.51)	2.68 (0.52)	2.64 (0.52)
Change from initial, mean (SD)	–	−0.02 (0.48)	−0.02 (0.49)	−0.03 (0.49)
P value	–	0.33	0.21	0.06
Effect size, SMG	–	−0.03	−0.04	−0.07

* All changes are calculated using only patients who had measurements at both points. EQ-5D-3L, EuroQol 5-Dimension 3-Level; HRQoL, health-related quality of life; MCS, mental component score; PCS, physical component score; SMG, standard mean gain; VR-6D, Veterans RAND 6-Dimension; VR-12, Veterans RAND 12-item.

of tophi (EQ-5D-3L), higher baseline serum urate, and higher RDCI score (Figure 2).

Mean ± SD HRQoL scores at baseline 24, 48, and 72 weeks are shown in Table 2. The mean overall VR-6D score at enrollment was 0.66 ± 0.12 with a baseline PCS of 36.8 ± 10.9 and an MCS of 51.3 ± 11.5. There were improvements in both the VR-6D and PCS by 24 weeks, which were sustained and significant at 72 weeks of follow-up (SMG 0.30 and 0.33 vs baseline, respectively, $P < 0.001$ for both). There were no significant changes in VR-12 MCS observed. Similar improvements in the EQ-5D-3L index score were observed, albeit of slightly lower magnitude (SMG 0.12, $P < 0.001$ at 72 weeks). Of the EQ-5D-3L dimensions assessed, only mobility (SMG 0.18) and pain (SMG 0.25) demonstrated evidence of improvement at 72 weeks ($P < 0.001$ for both) (Figure 3). There were no differences in the dimensions of self-care, activities, or mood at 72 weeks. In an additional analysis limited to participants with available HRQoL data from all study time points, changes in HRQoL over follow-up were nearly identical to those from our primary analysis that included data from all available individuals at each time point (Supplementary Table S3).

Results from GEE models examining factors associated with VR-6D and EQ-5D-3L changes at 24, 48, and 72 weeks from baseline are shown in Table 3. Improvements in both VR-6D and EQ-5D-3L index scores were associated with lower baseline hs-CRP (β −0.59 [95% CI −1.12 to −0.06] and

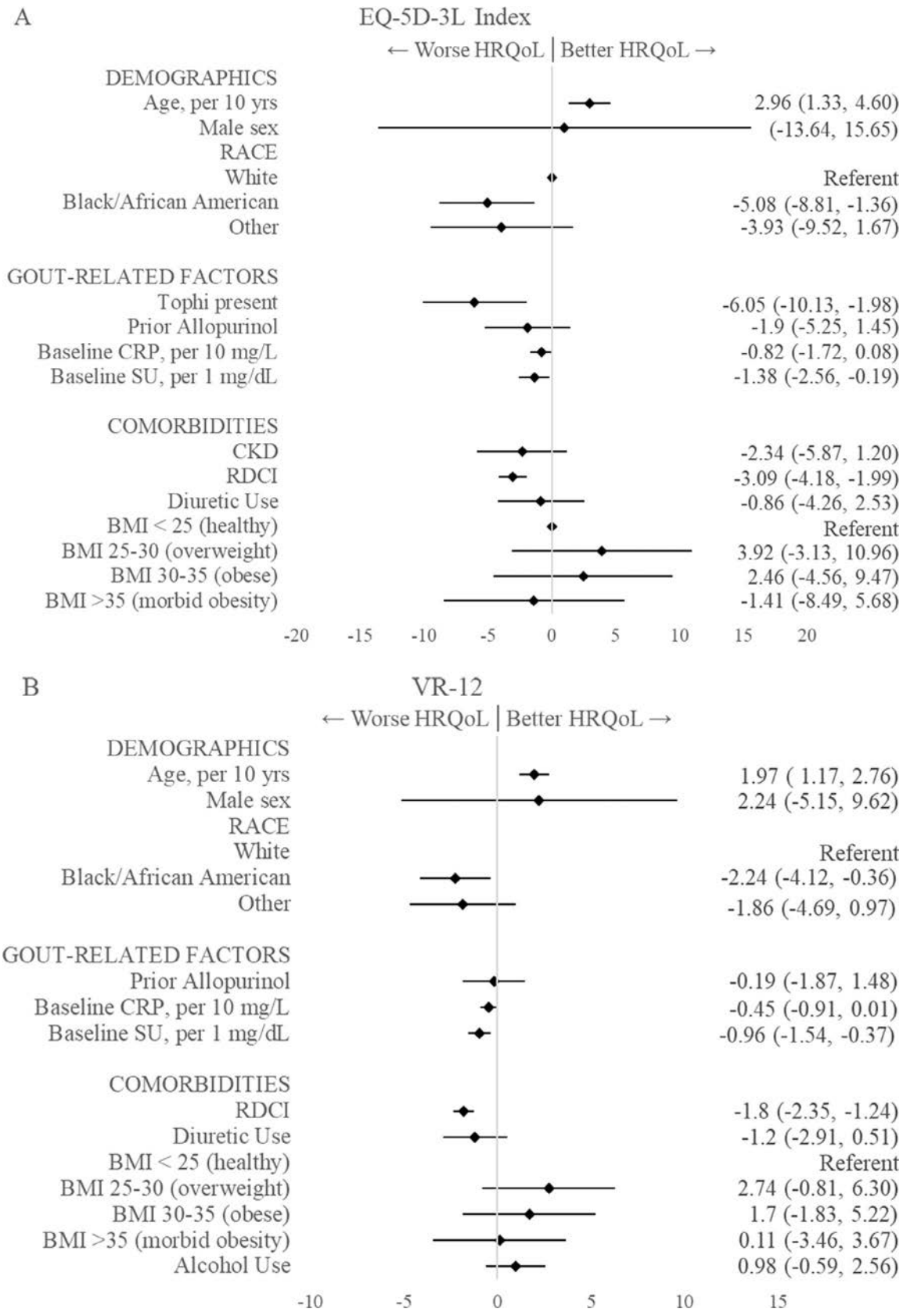


Figure 2. Legend on next page.

Figure 2. Multivariable linear regression of participant factors associated with baseline HRQoL in STOP Gout participants before initiating treat-to-target urate-lowering therapy. Beta coefficients and 95% confidence interval shown for (A) EQ-5D-3L Index and (B) VR-6D. Only factors with *P* value <0.1 in univariable analysis were included. EQ-5D-3L Index and VR-6D scores were multiplied by 100. Negative scores are associated with worse baseline HRQoL. "Other" race combines Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and "none of the above." BMI, body mass index; CKD, chronic kidney disease; CRP, C-reactive protein; EQ-5D-3L, EuroQol 5-Dimension 3-Level; HRQoL, health-related quality of life; RDCI, Rheumatic Disease Comorbidity Index; SU, serum urate; VR-12, Veterans RAND 12-item.

−1.20 [95% CI −2.29 to −0.11], respectively, per 10 g/L) and lower comorbidity burden as assessed by the RDCI (β −1.26 [95% CI −1.67 to −0.85] and −1.77 [95% CI −2.49 to −1.06], respectively). Consistent with similar outcomes noted in the parent trial,[13] there was no difference in HRQoL based on whether participants were randomized to allopurinol or febuxostat. Contrary to our a priori hypothesis, serum urate goal achievement at the end of the previous phase (observed in 83% at 24 weeks and 78% at 48 weeks) was not significantly associated with changes in HRQoL (Table 3). Although not achieving statistical significance, there were trends suggesting that those experiencing more frequent flares (at least two vs none; observed in 24% in phase 1 and 16% in phase 2) during the prior study phase were less likely to experience improvements in VR-6D (β −1.20; 95% CI −3.09 to 0.70) or EQ-5D-3L (β −0.86; 95% CI −1.86 to 0.13).

DISCUSSION

In this post hoc analysis using data from the STOP Gout trial,[13] participants receiving treat-to-target urate-lowering therapy demonstrated significant improvements in HRQoL over the 72-week study period. Improvements observed in both the VR-6D and EQ-5D-3L index scores were evident as early as 24 weeks and corresponded to small-to-moderate effect sizes. These improvements in HRQoL were driven largely by gains in the PCS of the VR-12 as well as in the pain and mobility components of the EQ-5D-3L. Although representing only small-to-moderate effects, changes observed in the VR-12 PCS approach or surpass the minimal important difference in this measure estimated for other musculoskeletal conditions.[28–30] These findings provide further support for treat-to-target urate-lowering therapy in gout management. In contrast, treat-to-target urate-lowering therapy was not accompanied by meaningful improvements in the MCS of the VR-12 or other HRQoL domains of the EQ-5D-3L (daily activities, mood, and self-care).

Results of this analysis complement and expand on those from other reports examining the impact of treat-to-target urate-lowering therapy on HRQoL.[10–12] A nurse-led trial in the United Kingdom showed improvements in HRQoL among participants with gout randomized to receive treat-to-target urate-lowering therapy,[12] improvements that were evident as early as 1 year following treatment initiation. In contrast, a parallel group of participants receiving only "usual gout care" demonstrated no meaningful changes in HRQoL over the same follow-up period. As in our analysis, improvements gained with treat-to-target urate-lowering therapy in the UK study were similar in magnitude and limited to the PCS (SMG 0.38) and did not extend to the MCS (measured in the UK study using the more comprehensive 36-item Short Form [SF-36] instrument). The earlier 6-month assessment available from the STOP Gout study suggests that gains in HRQoL are made sooner in the course of treat-to-target urate-lowering therapy than 1 year. The potential for earlier benefit in HRQoL is further supported by findings from the NOR-Gout trial, in which improvements in work participation and activity were evident within 3 months of treat-to-target urate-lowering therapy initiation.[10] Sustained benefit, observed in both the STOP Gout and NOR-Gout studies over 72 weeks and 2 years of follow-up, respectively, indicates that HRQoL improvements derived from treat-to-target urate-lowering therapy are likely independent of anti-inflammatory prophylaxis used earlier in treatment. Taken together, the results of these studies support both early and sustained benefit from treat-to-target urate-lowering therapy on HRQoL using first-line therapy (allopurinol or febuxostat) and the importance of treat-to-target urate-lowering therapy in improving patient-centered outcomes in gout.

In addition to demonstrating benefits accompanying treat-to-target urate-lowering therapy administration, data from the STOP Gout trial also provided an opportunity to identify factors associated with baseline HRQoL in participants with gout before initiating treat-to-target urate-lowering therapy as well as characteristics associated with improvements gained with best-practice management. In this post hoc analysis, factors independently associated with worse HRQoL in participants at the time of trial enrollment included younger age, Black/African American race (vs White), presence of tophi, higher baseline serum urate, and a higher comorbidity burden. These results align with those from separate investigations demonstrating associations of non-White race, gout severity, and gout-related comorbidities with reductions in HRQoL.[2,4,31–36] Previously, results from the STOP Gout study also showed that younger participants and those self-reporting non-White race were also less likely to achieve serum urate goals[17] whereas younger patients were more likely to experience arthritis flares complicating the course of treat-to-target urate-lowering therapy.[37] Alternatively, a younger age of gout onset has been proposed as an indicator of a more severe disease phenotype. In addition to being more recalcitrant to urate-lowering therapy, earlier-onset gout has been associated with

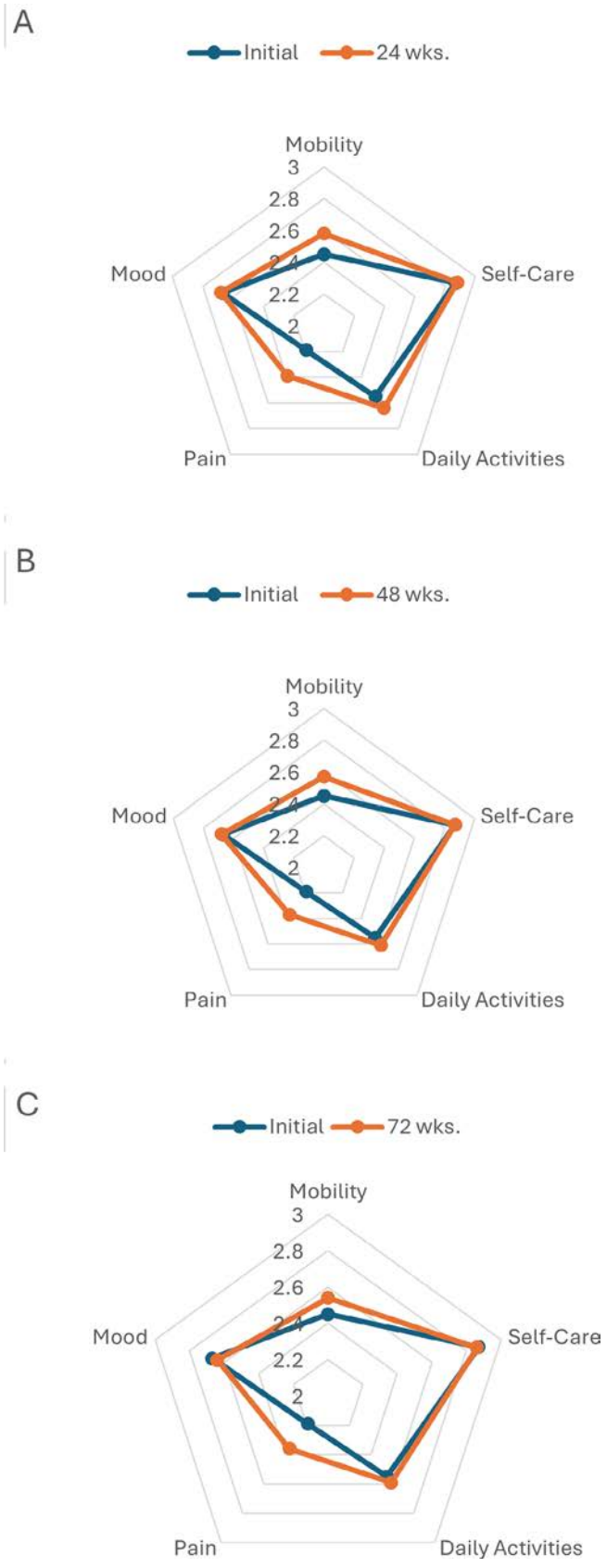


Figure 3. Legend on next page.

higher pretreatment serum urate level and a greater flare burden.[38] Associations of younger age with worse outcomes in gout contrasts with observations from the general population wherein HRQoL generally decreases with advancing age.[39] Although the reasons underpinning these different results across populations remain unknown, one possibility is that age-related differences in genetic factors influencing both disease susceptibility and outcome, drug metabolism, or changes in xanthine oxidase expression render younger patients less likely to benefit from treatment.

Beyond ascertaining gout-related determinants of HRQoL at enrollment, the availability of longitudinal trial data enabled the identification of factors associated with HRQoL improvements accompanying highly effective urate-lowering therapy. To date, there has been minimal evaluation of the factors associated with change in HRQoL produced within a treat-to-target framework. The results of the GEE models showed that improvements in HRQoL, whether measured by the VR-6D or EQ-5D-3L, were associated with a lower baseline comorbidity burden (lower RDCI score) and lower baseline levels of systemic inflammation as assessed by hs-CRP. The associations of comorbidity noted herein parallel findings from the NOR-Gout trial, demonstrating that fewer comorbidities were associated with greater improvements in work participation in the wake of treat-to-target gout management.[10] The reduced treatment response apparent among those with greater comorbidity may simply reflect QoL deficits that are independent of gout and the gout management strategies deployed.

In contrast to our a priori hypotheses, we did not detect associations of gout flares or serum urate goal achievement with HRQoL change. Although this study afforded a large sample size of participants receiving treat-to-target urate-lowering therapy, we had limited study power to address these post hoc questions, particularly as approximately 80% achieved the serum urate goal during the parent trial and >90% achieved serum urate concentrations <6.8 mg/dL.[13] Moreover, we were unable to assess flare severity in this analysis (eg, number of joints involved or pain intensity), which could be more closely related to changes in HRQoL than simple flare occurrence or count. Consistent with the noninferiority of allopurinol to febuxostat for the primary and secondary outcomes of flare reduction and serum urate goal achievement previously reported,[13] urate-lowering therapy assignment did not meaningfully impact HRQoL change.

Figure 3. Radar plots depicting change in the five dimensions (mobility, self-care, daily activities, pain, and mood) of the EuroQol 5-Dimension 3-Level from baseline to (A) 24 weeks, (B) 48 weeks, and (C) 72 weeks. Data were available for 877 participants at baseline, 748 at 24 weeks, 761 at 48 weeks, and 692 at 72 weeks.

Table 3. General estimating equation models examining factors associated with EQ-5D-3L and VR-12 changes at 24, 48, and 72 weeks versus baseline*

	EQ-5D-3L index		VR-6D	
	β (95% CI)	*P* value	**B** (95% CI)	*P* value
EQ-5D index baseline	−0.43 (−0.49 to −0.38)	<0.001	–	–
VR-12 baseline	–	–	−0.37 (−0.42 to −0.31)	<0.001
Demographics				
Age per 10 y	−0.41 (−1.48 to 0.65)	0.45	0.09 (−0.49 to 0.66)	0.77
Male sex	−0.28 (−5.95 to 5.39)	0.92	1.50 (−1.85 to 4.86)	0.38
Race				
White	Ref	Ref.	Ref	Ref.
Black/African American	−0.54 (−3.37 to 2.28)	0.71	0.33 (−1.36, 2.02)	0.70
Other	0.64 (−3.65 to 4.94)	0.77	0.80 (−1.71 to 3.31)	0.53
Hispanic ethnicity	1.62 (−2.94 to 6.19)	0.49	−0.28 (−2.96 to 2.40)	0.84
Urban residence	−1.87 (−4.12 to 0.39)	0.10	−0.91 (−2.28 to 0.45)	0.19
Gout-related factors				
Gout duration, per 10 y	−0.65 (−1.80 to 0.50)	0.27	−0.51 (−1.12 to 0.11)	0.11
Tophi present	1.88 (−0.92 to 4.67)	0.19	1.40 (−0.24 to 3.03)	0.09
Prior allopurinol use	−1.08 (−3.21 to 1.06)	0.32	0.07 (−1.21 to 1.34)	0.92
Baseline serum urate per 1 mg/dL	−0.59 (−1.36, 0.18)	0.13	0.07 (−0.37 to 0.50)	0.77
Treatment assigned				
Allopurinol	Ref	Ref.	Ref	Ref.
Febuxostat	0.59 (−1.42 to 2.59)	0.57	0.17 (−1.01 to 1.36)	0.77
Prophylaxis with colchicine	−1.58 (−4.97 to 1.81)	0.36	−0.89 (−2.83 to 1.05)	0.37
Flares in preceding phase				
0 flares	Ref	Ref.	Ref	Ref.
1 flare	−0.33 (−2.13 to 1.47)	0.72	0.16 (−0.75 to 1.08)	0.73
≥2 flares	−1.20 (−3.09 to 0.70)	0.22	−0.86 (−1.86 to 0.13)	0.09
Serum urate level at goal (end of prior phase)	0.52 (−1.35 to 2.38)	0.59	0.44 (−0.62 to 1.51)	0.41
Baseline hs-CRP per 10 mg/L	−1.20 (−2.29 to −0.11)	0.03	−0.59 (−1.12 to 0.06)	0.03
Comorbidities				
CKD Stage 3 (vs stage 1/2)	0.30 (−2.01 to 2.60)	0.80	0.23 (−1.09 to 1.54)	0.74
RDCI	−1.77 (−2.49 to −1.06)	<0.001	−1.26 (−1.67 to −0.85)	<0.001
BMI (kg/m^2)				
<25 (healthy)	Ref	Ref.	Ref	Ref.
25–30 (overweight)	0.15 (−3.48 to 3.79)	0.94	0.17 (−2.35 to 2.70)	0.89
30–35 (obese)	−0.33 (−4.06 to 3.40)	0.86	0.82 (−1.75 to 3.39)	0.53
>35 (morbid obesity)	−2.43 (−6.24 to 1.37)	0.21	−0.55 (−3.16 to 2.05)	0.68
Alcohol use	−1.22 (−3.27 to 0.83)	0.24	−0.81 (−2.03 to 0.40)	0.19

* "Other" race combines Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and "none of the above." BMI, body mass index; CI, confidence interval; CKD, chronic kidney disease; EQ-5D-3L, EuroQol 5-Dimension 3-Level; hs-CRP, high-sensitivity C-reactive protein; RDCI, Rheumatic Disease Comorbidity Index; Ref, reference; VR-6D, Veterans RAND 6-Dimension;VR-12, Veterans RAND 12-item.

There are limitations to this study. Although the US veteran population examined in this analysis demonstrates similar characteristics as those reported in other gout trials,[40–42] the results herein may not be universally generalizable. This issue might be most relevant among women, who made up <2% of STOP Gout participants.[13] Given similar findings of efficacy in the parent trial by treatment assignment, data from the original allopurinol and febuxostat arms were pooled for these analyses and all participants were administered urate-lowering therapy following a treat-to-target protocol. In the absence of an untreated or "usual care" comparator population, it is possible that improvements observed in HRQoL observed could attribute to other components of trial participation beyond urate-lowering therapy. Although the STOP Gout trial served as the source of data for this analysis, the post hoc study reported herein does not benefit from the initial randomization and has many of the limitations inherent to an observational study, including the possibility of participation bias or changes observed in HRQoL representing "regression to the mean."[43] Although impacting only a limited proportion of participants, it is also possible that study attrition could have biased findings from this analysis because individuals experiencing fewer benefits from the urate-lowering therapy intervention, including HRQoL improvements, may have also been more likely to have withdrawn from the study. This possibility is highlighted by differences observed in baseline EQ-5D-3L scores and residence of participants completing the study through 72 weeks and those dropping out. However, an analysis of participants with HRQoL data available at all time points closely mirrored that of the overall

study population, suggesting that attrition had minimal impact on these results.

In contrast to more comprehensive assessments used in other gout studies such as the SF-36,[11,12] the VR-12 and EQ-5D-3L instruments are brief, both requiring one-third of the time or less to complete. Although these generic health surveys may provide for greater study efficiency, the lower number of options for each item can result in a ceiling effect, defined as the inability of a questionnaire to capture increases in the measured attribute if participants already score at or near the maximum value.[39] Given the generic nature of the questionnaires used, it is possible that the use of a gout-specific instrument such as the Gout Impact Scale could have shown greater gains in HRQoL accompanying treat-to-target management.[44] However, we would anticipate that the generic nature of the questionnaires and any ceiling effect would both bias results toward the null, suggesting that our findings represent conservative estimates of HRQoL changes attending treat-to-target urate-lowering therapy.

In conclusion, by leveraging data available from a large comparative trial of allopurinol and febuxostat, this analysis demonstrates that treat-to-target urate-lowering therapy is accompanied by significant improvements in HRQoL. These improvements are most prominent in the domains of physical function, pain, and mobility; are evident relatively early in the course of treat-to-target therapy; and are sustained over longer periods of follow-up. Moreover, gout subgroups characterized by lower comorbidity burden and lower levels of systemic inflammation appear to experience the greatest improvements in HRQoL with treat-to-target urate-lowering therapy.

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 Mikuls 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. Yokose C, McCormick N, Lu N, et al. Trends in prevalence of gout among US Asian adults, 2011-2018. JAMA Netw Open 2023;6(4):e239501.
2. Chandratre P, Roddy E, Clarson L, et al. Health-related quality of life in gout: a systematic review. Rheumatology (Oxford) 2013;52(11):2031–2040.
3. Edwards NL, Sundy JS, Forsythe A, et al. Work productivity loss due to flares in patients with chronic gout refractory to conventional therapy. J Med Econ 2011;14(1):10–15.
4. Khanna PP, Perez-Ruiz F, Maranian P, et al. Long-term therapy for chronic gout results in clinically important improvements in the health-related quality of life: Short Form-36 is responsive to change in chronic gout. Rheumatology (Oxford) 2011;50(4):740–745.
5. Sigurdardottir V, Drivelegka P, Svärd A, et al. Work disability in gout: a population-based case-control study. Ann Rheum Dis 2018;77(3):399–404.
6. Scirè CA, Manara M, Cimmino MA, et al; KING Study Collaborators. Gout impacts on function and health-related quality of life beyond associated risk factors and medical conditions: results from the KING observational study of the Italian Society for Rheumatology (SIR). Arthritis Res Ther 2013;15(5):R101.
7. Schumacher HR, Taylor W, Edwards L, et al. Outcome domains for studies of acute and chronic gout. J Rheumatol 2009;36(10):2342–2345.
8. FitzGerald JD, Dalbeth N, Mikuls T, et al. 2020 American College of Rheumatology guideline for the management of gout. Arthritis Rheumatol 2020;72(6):879–895.
9. Richette P, Doherty M, Pascual E, et al. 2016 updated EULAR evidence-based recommendations for the management of gout. Ann Rheum Dis 2017;76(1):29–42.
10. Uhlig T, Karoliussen LF, Sexton J, et al. Course and predictors of work productivity in gout - results from the NOR-Gout longitudinal 2-year treat-to-target study. Rheumatology (Oxford) 2023;62(12):3886–3892.
11. Strand V, Khanna D, Singh JA, et al. Improved health-related quality of life and physical function in patients with refractory chronic gout following treatment with pegloticase: evidence from phase III randomized controlled trials. J Rheumatol 2012;39(7):1450–1457.
12. Doherty M, Jenkins W, Richardson H, et al. Efficacy and cost-effectiveness of nurse-led care involving education and engagement of patients and a treat-to-target urate-lowering strategy versus usual care for gout: a randomised controlled trial. Lancet 2018;392(10156):1403–1412.
13. O'Dell JR, Brophy MT, Pillinger MH, et al. Comparative effectiveness of allopurinol and febuxostat in gout management. NEJM Evid 2022;1(3).
14. Timilsina S, Brittan K, O'Dell JR, et al. Design and rationale for the Veterans Affairs "Cooperative Study Program 594 Comparative Effectiveness in Gout: allopurinol vs. Febuxostat" trial. Contemp Clin Trials 2018;68:102–108.
15. Khanna D, Khanna PP, Fitzgerald JD, et al; American College of Rheumatology. 2012 American College of Rheumatology guidelines for management of gout. Part 2: therapy and antiinflammatory prophylaxis of acute gouty arthritis. Arthritis Care Res (Hoboken) 2012;64(10):1447–1461.
16. US Food and Drug Administration. FDA adds boxed warning for increased risk of death with gout medicine Uloric (febuxostat). February 21, 2019. Accessed June 9, 2025. https://www.fda.gov/drugs/drug-safety-and-availability/fda-adds-boxed-warning-increased-risk-death-gout-medicine-uloric-febuxostat
17. Helget LN, O'Dell JR, Newcomb JA, et al. Determinants of achieving serum urate goal with treat-to-target urate-lowering therapy in gout. Arthritis Rheumatol 2024;76(4):638–646.
18. Kazis LE, Miller DR, Clark J, et al. Health-related quality of life in patients served by the Department of Veterans Affairs: results from the Veterans Health Study. Arch Intern Med 1998;158(6):626–632.
19. Kazis LE, Ren XS, Lee A, et al. Health status in VA patients: results from the Veterans Health Study. Am J Med Qual 1999;14(1):28–38.
20. Kazis LE, Lee A, Spiro A III, et al. Measurement comparisons of the medical outcomes study and veterans SF-36 health survey. Health Care Financ Rev 2004;25(4):43–58.

21. Jones D, Kazis L, Lee A, et al. Health status assessments using the veterans SF-12 and SF-36: methods for evaluating otucomes in the Veterans Health Administration. J Ambul Care Manage 2001; 24(3):68–86.
22. Kazis LE, Miller DR, Clark JA, et al. Improving the response choices on the veterans SF-36 health survey role functioning scales: results from the Veterans Health Study. J Ambul Care Manage 2004;27(3): 263–280.
23. Kazis LE, Miller DR, Skinner KM, et al. Patient-reported measures of health: the Veterans Health Study. J Ambul Care Manage 2004; 27(1):70–83.
24. Selim AJ, Rogers W, Qian SX, et al. A preference-based measure of health: the VR-6D derived from the Veterans RAND 12-Item Health Survey. Qual Life Res 2011;20(8):1337–1347.
25. England BR, Sayles H, Mikuls TR, et al. Validation of the Rheumatic Disease Comorbidity Index. Arthritis Care Res (Hoboken) 2015; 67(6):865–872.
26. Saary MJ. Radar plots: a useful way for presenting multivariate health care data. J Clin Epidemiol 2008;61(4):311–317.
27. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Routledge; 2013.
28. Zhou L, Natarajan M, Miller BS, et al. Establishing minimal important differences for the VR-12 and SANE scores in patients following treatment of rotator cuff tears. Orthop J Sports Med 2018;6(7): 2325967118782159.
29. Díaz-Arribas MJ, Fernández-Serrano M, Royuela A, et al. Minimal clinically important difference in quality of life for patients with low back pain. Spine 2017;42(24):1908–1916.
30. Sekhon S, Pope J, Baron M; Canadian Scleroderma Research Group. The minimally important difference in clinical practice for patient-centered outcomes including health assessment questionnaire, fatigue, pain, sleep, global visual analog scale, and SF-36 in scleroderma. J Rheumatol 2010;37(3):591–598.
31. Singh JA, Bharat A, Khanna D, et al. Racial differences in health-related quality of life and functional ability in patients with gout. Rheumatology (Oxford) 2017;56(1):103–112.
32. Topless R, Noorbaloochi S, Merriman TR, et al. Change in serum urate level with urate-lowering therapy initiation associates in the immediate term with patient-reported outcomes in people with gout. Semin Arthritis Rheum 2022;56:152057.
33. Becker MA, Schumacher HR, Benjamin KL, et al; Gout Natural History Study Group. Quality of life and disability in patients with treatment-failure gout. J Rheumatol 2009;36(5):1041–1048.
34. Hirsch JD, Terkeltaub R, Khanna D, et al. Gout disease-specific quality of life and the association with gout characteristics. Patient Relat Outcome Meas 2010;2010:1–8.
35. Singh JA, Sarkin A, Shieh M, et al. Health care utilization in patients with gout. Semin Arthritis Rheum 2011;40(6):501–511.
36. Singh JA, Strand V. Gout is associated with more comorbidities, poorer health-related quality of life and higher healthcare utilisation in US veterans. Ann Rheum Dis 2008;67(9):1310–1316.
37. Barry A, Helget LN, Androsenko M, et al. Comparison of gout flares with the initiation of treat-to-target allopurinol and febuxostat: a post-hoc analysis of a randomized multicenter trial. Arthritis Rheumatol 2024;76(10):1552–1559.
38. Amatucci AJ, Padnick-Silver L, LaMoreaux B, et al. Comparison between early-onset and common gout: a systematic literature review. Rheumatol Ther 2023;10(4):809–823.
39. McCaffrey N, Kaambwa B, Currow DC, et al. Health-related quality of life measured using the EQ-5D-5L: South Australian population norms. Health Qual Life Outcomes 2016;14(1):133.
40. Becker MA, Schumacher HR Jr, Wortmann RL, et al. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. N Engl J Med 2005;353(23):2450–2461.
41. Becker MA, Schumacher HR, Espinoza LR, et al. The urate-lowering efficacy and safety of febuxostat in the treatment of the hyperuricemia of gout: the CONFIRMS trial. Arthritis Res Ther 2010;12(2):R63.
42. Schumacher HR Jr, Becker MA, Wortmann RL, et al. Effects of febuxostat versus allopurinol and placebo in reducing serum urate in subjects with hyperuricemia and gout: a 28-week, phase III, randomized, double-blind, parallel-group trial. Arthritis Rheum 2008;59(11): 1540–1548.
43. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. Am J Epidemiol 2013; 177(4):292–298.
44. Wallace B, Khanna D, Aquino-Beaton C, et al. Performance of Gout Impact Scale in a longitudinal observational study of patients with gout. Rheumatology (Oxford) 2016;55(6):982–990.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 270–281
DOI 10.1002/acr.25612

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

A Scoping Review on Artificial Intelligence–Supported Interventions for Nonpharmacologic Management of Chronic Rheumatic Diseases

Nirali Shah,[1] Alexis Castellanos,[1] Yen Chen,[1] John Piette,[1] Amy Bucher,[2] and Susan Murphy[1]

This review summarizes artificial intelligence (AI)-supported nonpharmacological interventions for adults with chronic rheumatic diseases, detailing their components, purpose, and current evidence base. We searched Embase, PubMed, Cochrane, and Scopus databases for studies describing AI-supported interventions for adults with chronic rheumatic diseases. Eligible interventions targeted clinical outcomes (pain, function, disability, fatigue), psychological measures (depression, anxiety), or behavioral outcomes (physical activity, nutrition). All publication types (journal articles, conference abstracts, protocols) published in the English language until January 19, 2025, were considered, and interventions of any duration, frequency, country of origin, or setting (inpatient, outpatient, community, home setting) were included. Two reviewers independently screened studies, and one extracted data on study characteristics, intervention components, AI methodologies, and outcomes. Fifteen AI-supported interventions were identified, primarily targeting osteoarthritis (73%) and focusing on education and exercise advice (67%). The most common AI tool was rule-based expert systems (40%), followed by natural language processing systems (33%) and machine learning algorithms (27%). The interventions ranged from 3 weeks to 12 months, whereas sample sizes ranged from 7 to 427 participants, reflecting huge variability across studies. Most interventions demonstrated high usability, engagement, and adherence. Improvements in exercise compliance, physical activity, and symptoms, such as pain and physical function, were reported, though effects varied across studies and were sometimes not sustained long term. AI-supported interventions show promise in promoting education, exercise, and behavioral guidance for adults with chronic rheumatic diseases. There is evidence for high usability and engagement, but the clinical impact on long-term symptom management is uncertain.

Introduction

Rheumatic diseases, including osteoarthritis (OA), rheumatoid arthritis (RA), gout, systemic sclerosis (SSc), and systemic lupus erythematosus (SLE), are inflammatory, autoimmune, and multifactorial conditions that impose significant burdens on individuals, society, and health systems.[1–5] These diseases often lead to persistent pain, inflammation, and functional impairments in bones, muscles, and soft tissues, with some also affecting internal organs, such as the heart, lungs, or kidneys.[6] Rheumatic diseases frequently restrict mobility and daily activities and contribute to psychological distress, social isolation, and reduced quality of life.[7,8] Moreover, autoimmune rheumatic diseases, such as SSc and SLE, present with heterogenous and unpredictable clinical manifestation and often require a multidisciplinary and tailored approach to care.[9,10] Although pharmacological treatments can alleviate symptoms and slow disease progression, their long-term effectiveness varies, and many patients require additional nonpharmacological support for symptom management.[11,12] Nonpharmacological treatments, including behavioral and self-management interventions for improving physical activity, muscle strength, nutrition, sleep, and psychological beliefs, play a crucial role in management of rheumatic diseases by targeting symptom relief, functional improvement, and overall well-being.[13–16]

[1]Nirali Shah, PhD, PT, Alexis Castellanos, MS, Yen T. Chen, PhD, John D. Piette, PhD, MPH, Susan L. Murphy, ScD, OTR, FAOTA: University of Michigan, Ann Arbor; [2]Amy Bucher, PhD: Lirio, Inc, Behavioral Reinforcement Learning Lab, Knoxville, Tennessee.

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.25612).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25612.

Address correspondence via email to Nirali Shah, PhD, PT, at niralish@med.umich.edu.

Submitted for publication February 13, 2025; accepted in revised form July 3, 2025.

However, these interventions are often resource intensive and require repeated visits with trained specialists, thus limiting long-term adherence and accessibility.[17] As the global prevalence of rheumatic diseases continues to rise,[18] there is an increasing need for sustainable, scalable, and cost-effective approaches to personalized nonpharmacological management of rheumatic diseases.

Artificial intelligence (AI) refers to a broad set of technologies that simulate human learning, comprehension, problem-solving, decision-making, and creativity.[19] With advancements in AI technologies, researchers have increasingly explored its applications for nonpharmacological management of chronic conditions, including rheumatic diseases.[20–23] These AI-supported interventions use different approaches, ranging from expert systems that use predefined rules to provide treatment recommendations to more adaptive machine learning algorithms that learn patterns from large datasets to provide more personalized treatment recommendations.[24,25] Some interventions leverage subtypes of machine learning, such as deep learning and reinforcement learning, which allows continuous adjustment of treatment recommendations based on objective data or real-time patient feedback.[26,27] Finally, AI tools that enable understanding, interpretation, and generation of human language, such as large language models and natural language processing (NLP), have been used to provide human-like health coaching experience to patients to help them manage their behavioral goals.[28] Together, these studies show the potential of AI tools to improve efficiency, accessibility, and personalization of nonpharmacological management approaches for chronic rheumatic diseases.

The rise of AI marks a transformative moment that has the potential to shape human interaction and power structures.[29] This raises important questions about its added value and thoughtful integration within existing health care systems for rheumatic disease management. However, the novelty and interdisciplinary nature of AI-supported interventions make it challenging for clinicians and researchers to navigate this emerging field.[30] A comprehensive understanding of existing applications of AI and evidence gaps is needed to inform future research and clinical translation.[31] Although prior reviews on the application of AI for management of chronic rheumatic diseases focused on improving diagnosis, prognosis, and disease phenotyping,[32–34] a review specifically focused on nonpharmacological management of chronic rheumatic diseases is lacking in the literature. Addressing this gap will improve our understanding of AI's potential for nonpharmacological management, identify areas for further exploration, and guide the development of effective and innovative AI-supported interventions for managing rheumatic diseases. Therefore, this scoping review summarizes the evidence on AI-supported nonpharmacological interventions for individuals with chronic rheumatic diseases.

Methods

Search strategy and terms. The protocol for this review was registered a priori on the Open Science Framework (osf.io/8b6cw). We used the Arksey and O'Malley framework to guide our research approach.[35] Adhering to the process outlined in the framework, we conducted this review by first defining the research question, identifying relevant studies, charting data, and collating and summarizing studies to get an understanding of the current evidence.[35,36] We have reported this review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines (Figure 1, Supplementary Material 1).[37]

We consulted with the library informationist at the University of Michigan to develop our search strategy and terms. We searched for studies published until January 19, 2025, using the Embase, PubMed, Cochrane, and Scopus databases. These databases were selected because they cover relevant studies in the field of digital health, rehabilitation, and self-management. The search terms included predefined keywords and medical subject heading terms related to digital health, AI, interventions, symptom management, behavior change, and chronic rheumatic diseases. These search terms were developed to answer our overarching research question, "What is the evidence supporting the use of AI for nonpharmacological management, that is, interventions designed to improve symptoms and change behavior, for people with chronic rheumatic diseases?" We included all publication types (conference abstracts, journal articles, and research protocols) to ensure a comprehensive coverage of literature. Additionally, reference lists of relevant articles were examined for any studies potentially meeting the inclusion criteria. A detailed description of our search strategy and terms for all databases are provided in Supplementary Material 2.

Inclusion and exclusion criteria. We applied the population, intervention, comparator, outcome (PICO) framework to define our search criteria.[38] For population, we included adults with chronic rheumatic diseases, as defined by the National Institute of Arthritis and Musculoskeletal and Skin Diseases.[6] These conditions include ankylosing spondylitis, autoimmune rheumatic diseases, gout, polymyalgia rheumatica, giant cell arteritis, SSc, Sjögren disease, SLE, and various forms of arthritis such as OA, RA, psoriatic arthritis, reactive arthritis, as well as indiviudals who have undergone total hip and knee replacements. For intervention, we included digital health interventions that were designed to improve clinical outcomes (pain, function, disability, fatigue), psychological measures (depression, anxiety), or behavioral outcomes (physical activity, nutrition, adherence to treatment). Interventions conducted in any setting (inpatient, outpatient, community, home) and of any duration or frequency were eligible for inclusion. Although studies from any country of origin were eligible for inclusion, only articles published in English were

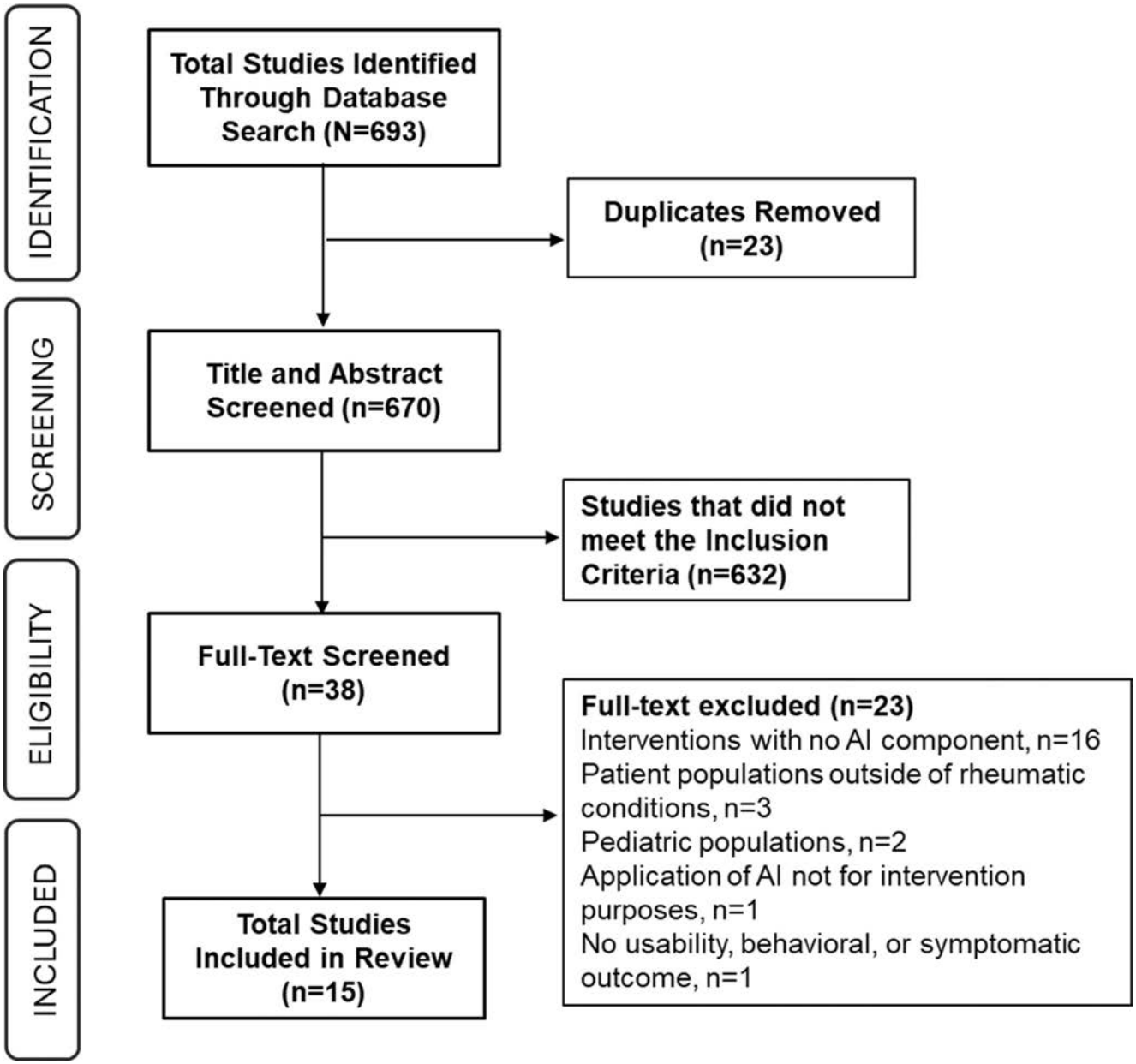


Figure 1. PRISMA-ScR flow diagram of identification, screening, and selection of review articles. AI, artificial intelligence; PRISMA-ScR, Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews.

considered. For comparator, we included studies of all designs, including single-arm studies and randomized controlled trials with or without a comparison group. Finally, for outcome, we included studies that examined usability, symptom management, or behavior change. The detailed eligibility criteria are outlined in Table 1.

Article selection. All articles identified through the search were uploaded on the Covidence Review Software.[39] The article selection and screening was conducted by two reviewers (NS and YTC) in two phases. In the first phase, the reviewers independently screened all titles and abstracts of the articles to determine eligibility of studies based on predefined inclusion and exclusion criteria. Any disagreements between the reviewers during this stage were resolved through discussion, and any misunderstandings regarding the inclusion criteria were clarified to ensure consistent application throughout the remainder of the evaluation process. In the second phase, the full text of the included articles was reviewed to ensure that they met the inclusion criteria. During this phase, interventions were assessed to confirm that they incorporated an AI component. Interventions were considered AI-supported if they provided recommendations or monitored participants through decision-making, recognizing patterns, or adapting based on new information.[19] This included

Table 1. Inclusion and exclusion criteria for studies included in the review, outlined by the PICO framework*

PICO framework	Inclusion criteria	Exclusion criteria
Patient/study population	Adults with chronic rheumatic diseases, as defined by NIAMS. Based on this definition, the conditions included in the review are ankylosing spondylitis, arthritis, autoimmune diseases, autoinflammatory diseases, gout, juvenile idiopathic arthritis, OA, polymyalgia rheumatica and giant cell arteritis, psoriatic arthritis, reactive arthritis, RA, scleroderma, Sjögren disease, and SLE.	Animal studies or studies in adults who are healthy, pain free, or have acute injuries or conditions
Intervention	Digital health intervention, of any duration or frequency, with an AI component, designed for nonpharmacological management and conducted in any setting (inpatient, outpatient, community, and home setting) or country	Pharmacological or surgical treatment or digital health intervention without an AI component
Comparator	Any or none	–
Outcome	Any usability, clinical (pain, function, disability, fatigue), psychological (depression, anxiety), or behavioral (physical activity, nutrition) outcome	Outcomes related to prediction, diagnosis, or risk factor analysis

* AI, artificial intelligence; NIAMS, National Institute of Arthritis and Musculoskeletal and Skin Diseases; OA, osteoarthritis; PICO, population, intervention, comparator, outcome; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

expert systems, which use structured rule-based reasoning and large databases to analyze information and generate recommendations based on predefined logic in a manner intended to replicate aspects of clinical reasoning.[40] Additionally, interventions with a machine-learning or NLP component that learned from objective data or participant feedback to provide contextual recommendations were included.[41,42] In contrast, interventions that incorporated static automation, such as if/then rules without any reasoning, decision-making, or adaptation,[43] were excluded from this review.

Data extraction. Data extraction was performed independently by one reviewer (NS) using a structured, prespecified Microsoft Excel spreadsheet, which was developed based on the objectives of the review.[36] Extracted information included study characteristics, patient population details, intervention components, and study outcomes. For study characteristics, we recorded author details, year of publication, country where the research was conducted, type of publication (eg, journal article, conference abstract), and study design (eg, randomized controlled trial, pilot study). For patient population, we included the type of chronic rheumatic disease targeted, sample size, and demographic characteristics, such as age, sex, and ethnicity. For intervention components, we included the primary purpose of the intervention, the AI tool used, the delivery medium of intervention (eg, mobile app, web), intervention duration and frequency, and details on the comparison group if applicable. Finally, study outcomes related to usability, behavior, and symptoms were extracted.

Evidence synthesis. Following data extraction, interventions were categorized based on their primary purpose. The categories included education and exercise, gait retraining, clinical decision support, remote patient monitoring, and pain management. Within these, interventions were further subcategorized based on the AI approach used in the intervention. The subcategories on AI approaches were developed in consultation with a machine learning expert (AC) at the University of Michigan. The expert provided guidance on how to distinguish the AI categories and classify the interventions in the included studies appropriately. The AI approaches included the following: NLP systems, rule-based expert systems, recommender systems, and motion analysis systems. Interventions were classified as NLP-based if they used AI to process, understand, and generate human language.[42] Interventions were classified as expert systems if they included a structured knowledge base (eg, a list of predefined treatment options) and used AI to infer patient data, apply reasoning, and provide treatment recommendations based on if/then rules, similar to a human expert.[40(pX)] Interventions were classified as recommender systems if they used AI to analyze patient data and provide personalized treatment recommendations based on individual characteristics, preferences, or behavioral data.[44] Interventions were classified as motion analysis systems if they used machine learning algorithms to identify patterns from movement data to provide real-time feedback on exercise movements.[41] This included interventions that used computer vision (eg, camera of smartphone) or data from wearable sensors (eg, inertial movement units) for motion analysis and deep learning for pattern recognition in health data. The studies included in this review typically used multiple AI approaches to deliver their intervention.

To synthesize data, we first descriptively summarized study and participant characteristics (eg, country, condition targeted, sample size, study design, AI method used, outcomes). Then,

we conducted a narrative synthesis, organizing findings under the prespecified categories of intervention purpose and AI approach to highlight trends, effectiveness, and evidence gaps across studies.

Results

Search results. A total of 693 studies were identified through the database search, with 670 studies remaining after removal of duplicate entries. Following independent reviews by two researchers, 38 studies were selected for full-text review. Of these, 23 studies were excluded for not meeting the inclusion criteria, the most common reason for exclusion (~70%, 16 of 23) being that the automation in the interventions did not include an AI component. A total of 15 studies were included in this review (Figure 1).

Overview of included studies. Of the 15 included studies, a majority focused on individuals with OA (73%, 11 of 15), including two specifically designed for individuals undergoing total knee replacement.[25,26,45–54] Other populations included individuals with RA (n = 1), with autoimmune inflammatory rheumatic diseases (AIIRDs; n = 1), and with SLE (n = 1) and postmenopausal women with osteoporosis (n = 1).[55–58] Of these, eight studies reported the age of participants, which ranged from 48 to 67 years.[25,26,45,48–50,56,57] The sample size of participants varied significantly, ranging from 7 to 427 participants. Geographically, the studies were conducted across Europe (33%), the United States (27%), Asia (33%) and Australia (7%). Most interventions were evaluated in randomized controlled trials (73%, 11 of 15), with the remaining being evaluated in development and pilot trials (27%, 4 of 15). Interventions ranged from three weeks to six months and were primarily delivered using mobile apps (53%, 8 of 13).

Characteristics of interventions by purpose. The AI-supported nonpharmacological interventions for adults with chronic rheumatic diseases are described based on their primary purpose. An overview of the interventions is provided in Table 2, with outcomes and brief findings described in Supplementary Material 3.

About 67% (10 of 13) of interventions included in this review aimed to promote disease-related education and exercise for adults with OA, RA, and osteoporosis.[26,45,46,48,50–53,55,56] Of these, 50% of the interventions incorporated NLP systems,[46,51,53,56] followed by rule-based expert systems (40%)[26,45,48,50] and recommender systems (10%).[55] Among these, 40% of interventions integrated motion analysis systems to provide real-time feedback on exercise movement along with the treatment recommendations.[26,46,51,52]

AI approaches used for education and exercise. *NLP systems.* Two of the five interventions that incorporated NLP systems used a third-party large language model, OpenAI's ChatGPT, to provide home exercise regimens for people with knee OA and compared them with in-person physical therapy.[51,52] Along with exercises recommended by ChatGPT, participants were given real-time feedback on their exercise movements using either computer vision (camera system that captures motion) or wearable sensors (inertial measurement units).[51,52] In contrast, Blasco et al developed a NLP-based conversational agent for people with knee OA undergoing total knee replacement that provided daily automated messages with education, exercise recommendations with video instructions, progress reports, and motivational messages.[53] Feng et al developed a multimodal intervention that incorporated computer vision to capture and analyze exercise movement and a deep learning system powered by a large language model to describe the exercise movements, to ensure correct exercise form and adherence.[46] In another application of NLP systems, Scott et al used a digital voice assistant, Amazon Alexa Echo, to broadcast education and exercise, provided by an exercise physiologist.[56]

Among the five studies that used NLP-based systems for education and exercise advice, only two studies reported results, whereas the remaining were study protocols. Both studies found improvements in exercise adherence.[53,56] Blasco et al reported a 15% higher compliance rate among individuals who used their chatbot intervention compared to those receiving standard postoperative care at 12 weeks.[53] Similarly, Scott et al found high adherence in participants using Alexa, who completed an average of 68 out of 72 sessions.[56] Although improvements in moderate–vigorous physical activity (+17.9 ± 28.8 m/d) and sedentary behavior (–40.2 ± 71.7 m/d) were observed in the Alexa group, these changes were comparable to the control group receiving monthly educational emails.[56] Improvements in calcium intake and medication adherence attitudes were also similar across groups.[56] However, participants in the Alexa group showed significantly greater improvements in falls efficacy ($P = 0.032$) and osteoporosis knowledge ($P = 0.038$) compared to controls.[56]

Rule-based expert systems. The interventions in this subsection included predefined education and exercise libraries and used various factors to create tailored goals from the predefined list. For example, dr. Bart by Pelle et al analyzed user data and proposed three sets of preformulated goals from the intervention library.[59] The intervention continued to propose goals until three goals were chosen by the user, and after one or more goals were completed, a new set of goals were proposed.[59] Kaia Health proposed education, exercises, and relaxation techniques based on user needs along with real-time feedback on exercises using the front facing camera of a smartphone or tablet (computer vision).[26] In contrast, Rak Kao by Thiengwittayaporn et al analyzed questionnaire data to assess disease stage and generated treatment recommendations from a predefined list of treatment

Table 2. An overview of AI-supported interventions included in this review*

Study/ intervention	Population, N	Primary purpose	Intervention description	Medium	Duration	Type of AI
You et al, 2024[52]	OA and sarcopenia, 278	Exercise	Tailored exercise regimen generated by a large language model (ChatGPT4), real-time feedback on exercise movements	Chatbot, wearable sensors	3 mo	NLP
Xu et al, 2024[51]	Knee OA, 60	Exercise	Exercise regimen generated by a large language model (ChatGPT4), real-time feedback on exercise movements	Chatbot, computer vision	12 wk	NLP
Blasco et al, 2024[53]	People with knee OA undergoing TKR, 40	Education, exercise	Treatment recommendations, progress reports, motivational messages	Chatbot	N/A	NLP
Thiengwittayaporn et al, 2023[50] Rak Kao	Knee OA, 82	Education, exercise	Treatment recommendations based on disease stage	Mobile app	4 wk	Rule-based expert systems
Tan et al, 2023[57] RheumConnect	AIIRDs, 200	Clinical decision support system	Assistance with referrals, disease education, medication refills, and appointment scheduling	Chatbot	N/A	NLP
Scott et al, 2023[56]	Postmenopausal women with osteoporosis, 47	Education, exercise	Physiologist prescribed exercise broadcast through Amazon Alexa Echo	Digital voice assistant	24 wk	NLP
Sanchez et al, 2023[58]	SLE, 7	Clinical decision support system	Treatment recommendations based on disease stage and medication use	Website	N/A	Rule-based expert systems
Feng et al, 2023[46]	Knee OA, 86	Education, exercise, behavioral health coaching	Real-time feedback on exercises using wearable sensors and a large language model that describes movements	Mobile app, wearable sensors, video for health coaching	12 wk	Motion analysis, machine-learning algorithms, and NLP
Studenic et al, 2022[55] RheumaBuddy	RA, N/A	Education, exercise, behavioral health coaching, peer support, health monitoring	Treatment recommendations based on user's symptoms, mood, claims, and clinical data. Based on resulting behavior, algorithms created groups for peer support	Chatbot	N/A	Recommender systems, machine-learning algorithms
Bobrovsky et al, 2022[26] Kaia Hip and Knee Pain	Hip or knee OA, 192	Education, exercise	Real-time audiovisual feedback on exercises, education, and relaxation techniques	Mobile app, computer vision	90 d	Motion analysis, machine-learning algorithms

(Continued)

Table 2. (*Cont'd*)

Study/ intervention	Population, N	Primary purpose	Intervention description	Medium	Duration	Type of AI
He et al, 2021[47]	Knee OA, 38	Gait retraining	Real-time auditory feedback on gait pattern. User is instructed to shift weight medially when lateral pressure exceeds preset thresholds	Mobile app, pressure-based sensors in insole of shoe	3 wk	Motion analysis, machine-learning algorithms
Pelle et al, 2020[48] dr. Bart	Hip or knee OA, 427	Education, exercise	Exercise goals based on profile data and previously chosen and discarded goals	Mobile app	6 mo	Rule-based expert systems
Ramkumar et al, 2019[49] TKR	People with knee OA undergoing TKR, 25	Remote patient monitoring	Motion data are transmitted from the knee sleeve to the mobile app, and an avatar is used to depict knee ROM during exercises	Mobile app with motion sensors in knee sleeve	3 mo	Motion analysis, machine-learning algorithms
Rini et al, 2015[25] PainCOACH	Hip or knee OA, 113	Pain management	Tailored PCST modules based on user's experience and progress through online program	Website	8 wk	Rule-based expert systems
Bossen et al, 2013[45] Join2move	Hip or knee OA, 199	Education, exercise	Tailored modules based on a baseline physical activity test, short-term goals, and favorite recreational activity	Website	8 wk	Rule-based expert systems

* AI, artificial intelligence; AIIRD, autoimmune inflammatory rheumatic disease; N/A, not available; NLP, natural language processing; OA, osteoarthritis; PCST, pain coping skills training; RA, rheumatoid arthritis; ROM, range of motion; SLE, systemic lupus erythematosus; TKR, total knee replacement.

recommendations.[50] Although the prior studies considered user profile, needs, and disease-related information, Join2move by Bossen et al used behavioral data and personal preferences to create eight weekly tailored goals for the user.[45] This included a baseline physical activity test, user's short-term goals, and their favorite recreational activity from a list of predefined activities.[45]

Among the four studies that used rule-based expert systems for education and exercise, findings showed high adherence (defined as choosing at least one behavioral goal or completion of modules or prescribed exercises)[45,48,50] and exercise accuracy (defined as ability to correctly perform exercises).[50] Thiengwittayaporn et al reported that participants using the Rak Kao app completed 8 of 10 prescribed repetitions and showed improvements in symptoms, satisfaction, and quality of life compared to those who received handouts.[50] However, there were no between-group differences in range of motion or self-reported symptoms, as measured by the Knee Injury and Osteoarthritis Outcome Score (KOOS) and the Knee Society Score.[50,60,61]

The Join2move program by Bossen et al reported high initial engagement (94% of participants started the program), but only about half (46%) completed at least six of nine modules.[45] Despite this, the intervention group demonstrated greater self-reported and objective physical activity at 12 months compared to a waitlist control group.[45] Also, there were significant improvements in physical function and self-perceived benefit at 3 months, but they were not sustained at 12 months.[45] Similarly, Pelle et al reported high app engagement but modest usability (System Usability Scale [SUS][62] scores of 65.9 at three months and 64.5 at six months) and a weak association between app use and symptom outcomes.[48] Bobrovsky et al evaluated the Kaia Hip and Knee Pain app and observed clinically meaningful improvements in self-reported symptoms (Western Ontario and McMaster Universities Osteoarthritis Index [WOMAC] scores)[63] among the intervention group compared to usual care; however, no significant differences were observed in overall quality of life, health care use, and number of lost workdays.[26] Overall, although these interventions demonstrated potential for improving adherence and physical activity, their impact on long-term symptom management and usability was variable and often modest when compared to waitlist or usual care control conditions.

Recommender systems. Studenic et al evaluated a multimodal AI-supported intervention Rheumabuddy, which provided education and exercise advice, health coaching, peer support, and health monitoring.[55] The intervention used a recommender system to analyze user data, including symptoms, mood, and claims data to provide personalized treatment recommendations.[55] Additionally, it used machine learning algorithms to group users with similar symptoms and behavioral patterns, facilitating peer support.[55]

The findings of the implementation study were not reported at the time of this review. However, Rheumabuddy is currently used by more than 3,100 people with RA in 35 countries across Europe.[55]

AI for gait retraining. He et al developed a gait retraining intervention for individuals with knee OA using pressure-based auditory feedback.[47] The intervention used a motion analysis system that detected pressure on the insole of the shoe, transmitted the data to a mobile app, and analyzed it using machine learning algorithms.[47] When lateral pressure on the insole exceeded preset thresholds, the mobile app generated auditory cues and instructed the user to alter their gait by shifting their weight medially while walking.[47]

The gait retraining intervention resulted in high adherence and short-term improvements in gait and symptoms.[47] Participants who received the pressure-based auditory feedback showed significant reductions in knee adduction moment immediately post intervention compared to those who received gait monitoring without auditory feedback. Additionally, the intervention group reported greater improvements in pain (measured by visual analog scale [VAS])[64] and self-reported symptoms on KOOS[61] when compared to controls.[47] However, the between-group differences in outcomes were not retained at three- and six-week follow-ups, indicating that the effects of the intervention may diminish over time without continued feedback.[47]

AI for clinical decision support. Two studies focused on developing clinical decision support systems for individuals with SLE and AIIRDs, respectively.[57,58] Sanchez et al created the SLE-T2T clinical decision support system, which used a rule-based expert system to provide clinicians with treatment recommendations for SLE.[58] The recommendations were structured using an if/then decision table derived from established guidelines and considered factors such as disease activity (remission, mild, or moderate) and medication use (antimalarials, immunosuppressants, or glucocorticoids) in the decision-making process.[58] This system received a high usability score (SUS = 79 of 100) and acceptance from health care professionals, with suggestions to improve functionality for patients.[58] Tan et al developed the RheumConnect chatbot, which integrated NLP and rule-based expert systems.[57] This chatbot assessed user symptoms through questionnaires, referred patients to appropriate physician or specialist care, facilitated appointment scheduling, and supported medication refills.[57] The chatbot was developed for multi-ethnic populations and is currently used across hospitals in Singapore.[57]

AI for remote patient monitoring. Ramkumar et al developed a remote patient monitoring mobile application for people with knee OA undergoing total knee replacement.[49] The system included a motion analysis system that collected mobility data from motion sensors in a neoprene knee sleeve and transmitted it to the mobile app, which used machine learning algorithms to analyze and record the user's range of motion and step count. Additionally, it collected data related to the symptoms, opioid use, and adherence to exercise.[49]

This single-arm pilot study demonstrated feasibility of real-time tracking and analysis of mobility data to support postoperative recovery following total knee replacement.[49] Along with high engagement (inferred from continuous interrupted motion and symptom data) and adherence (defined by exercise compliance), participants reported significant improvements in step count (mean = 4,654 steps/d), range of motion (mean knee flexion = 119°), and self-reported symptoms (mean KOOS improvement = 39.3) at three months.[49] Moreover, participants stopped opioid use by postoperative day five.[49] Although there was no comparator group, the reported improvements were consistent with expected recovery trajectories seen in standard clinical care.[49]

Pain management. Rini et al used rule-based expert systems to create a web-based pain coping skills training program, PainCOACH.[25] The program created eight weekly modules for the user based on their pain experience and progress through the program.[25] Participants in this program were led though the program through a virtual coach, who provided verbal instructions, feedback, and encouragement through each module.[25]

Participants in the PainCOACH group showed high engagement, with 91% completing all eight web-based modules. Compared to a waitlist control group, individuals in the intervention group reported significant reduction in pain (Cohen's d = 0.33) and self-efficacy for pain management (Cohen's d = 0.43) at post-intervention.[25] Effects on other outcomes, such as pain-related interference, pain anxiety, and positive and negative affect, were not statistically significant.

Discussion

This review summarizes the current application of AI in the nonpharmacological management of chronic rheumatic diseases, identifying 15 AI-supported interventions designed for OA, RA, AIIRDs, SLE, and osteoporosis. Among these, the predominance of interventions focused on OA highlights the need to expand AI applications to rare rheumatic conditions, such as SSc, for which health care providers are scarce and the disease heterogeneity necessitates individualized care. Although most interventions were patient facing, some systems used AI for clinical decision-making or as digital diagnostic tools,[57,58] suggesting opportunities for AI integration not only in patient self-management but also in enhancing clinician workflow and treatment planning.

The majority of the interventions focused on disease education and exercise advice, suggesting a potential gap in addressing additional management needs, such as nutritional guidance and psychological support.[7,65] Although most studies demonstrated high usability, engagement, and exercise adherence, symptomatic improvements were often modest and, in some patients, comparable to control groups. The most commonly used AI approach used was rule-based recommender systems, highlighting an opportunity to diversify AI methodologies in rheumatology with models like reinforcement learning, which could optimize treatment personalization, or NLP-driven virtual health coaching for emotional and behavioral support. Overall, although there is clear potential for AI integration in nonpharmacological interventions for rheumatic diseases, further research is needed to establish robust evidence of efficacy, assess long-term outcomes, and expand AI applications beyond exercise and education-based interventions.

The AI-supported interventions included in this review primarily focused on education, exercise, and behavioral guidance by providing personalized treatment recommendations, tailored based on factors such as disease stage,[50] clinical data (eg, symptoms and mood),[55,59] and baseline physical activity and user preferences.[45] Several interventions used objective, sensor-based data to track and correct movements or gait,[26,46,49,51,52] adding a layer of real-time monitoring to improve exercise accuracy and tracking. Others used chatbots or digital voice assistants for education, motivation, exercise, and to assess compliance to exercise treatments.[46,53,56] Across studies, high levels of satisfaction, engagement, and exercise adherence were reported across studies, reflecting the acceptability of these AI-supported interventions among participants.

Although the studies in this review highlight the potential of using AI for chronic rheumatic disease management, many studies were in the early stages of development and evidence generation. Moreover, although several studies reported clinically meaningful improvements in self-reported symptoms, function, and physical activity, these improvements were often modest, and in some cases, not significantly different from control groups (usual care, handouts, waitlist, or monthly educational emails).[25,26,45,47–49,56,59] Notably, mobile-based self-management interventions without an AI component have demonstrated comparable levels of engagement and symptom improvement in the literature.[66,67] As no studies in this review directly compared an AI-supported intervention with a non-AI digital intervention, it is unclear whether AI-supported approaches offer significant advantages over non-AI interventions. As such, further research is needed to evaluate the added value of AI components and determine if they lead to meaningful clinical benefits over traditional self-management interventions.

Other AI applications in the nonpharmacological management of rheumatic diseases included a chatbot designed to assist with referrals based on patient-reported symptoms,[57] a recommender system that identifies treatment plans through patient-reported symptoms,[68] and a web-based pain management program using a traditional expert system approach.[25] Although these interventions represent innovative advancements, future AI-supported interventions should consider extending their focus to additional aspects of disease management, such as medication adherence, lifestyle modifications (eg, diet, sleep), psychological support, and long-term symptom tracking, which are essential components of comprehensive care for rheumatic

diseases. Moreover, further research is needed to assess the long-term clinical benefits and usability of such systems across diverse patient populations.

Many studies included in this review lacked diversity in participant populations, limiting generalizability across different demographic or cultural groups. Yet, AI may be well-suited to tailoring for diverse populations. A notable example of culturally tailored AI is the referral chatbot developed for Singapore's multi-ethnic Asian population.[57] By adapting recommendations based on nuanced cultural understanding, this chatbot aligns more effectively with the diverse patient demographics in Singapore, potentially offering more relatable and acceptable responses in multi-ethnic settings. This highlights the value of developing AI chatbots for disease management that incorporate cultural and contextual nuances to enhance patient engagement and improve health outcomes across various populations. Another example of a multilingual AI-supported system is Rheumabuddy, which provides behaviorally focused recommendations based on patients' clinical data and integrates claims information to offer personalized treatment options and peer support.[55] This intervention also groups individuals with similar symptom and behavior patterns, fostering community support, and is currently available in eight languages across 35 countries in Europe. These interventions demonstrate the potential of AI to provide tailored, accessible support for patients across diverse cultural and health care contexts. Thus, expanding AI applications to encompass broader aspects of disease management while customizing these systems for culturally diverse populations may significantly enhance the relevance and effectiveness of AI-supported interventions in rheumatic disease management.

A key limitation in the literature on AI-supported interventions for rheumatic diseases is the lack of a consistent definition of what constitutes AI. Although some researchers exclude expert systems from the definition,[22] we included these systems to capture a broader range of AI technologies used in interventions. This inconsistency in terminology highlights the need for greater clarity in defining AI across studies. Furthermore, many studies refer to "artificial intelligence" or "machine learning algorithms" without specifying the exact tools or frameworks used, which complicates assessments of effectiveness and limits replicability. The reliance on proprietary AI systems further restricts comparability and reproducibility across studies. This challenge has been previously noted in AI research,[22] and it emphasizes the need for transparency in reporting the detailed descriptions of the AI models, data inputs, and any customization used in interventions.

Many studies focus on feasibility or pilot outcomes without clearly defining the expected impact or identifying pathways for AI to improve patient care. Given that these interventions incorporated different factors to personalize advice and behavioral guidance, it remains unclear which elements, such as disease severity, user preferences, clinical data, or a combination of these, are most effective in achieving sustained improvements in adherence, physical function, and overall outcomes. Furthermore, interventions that used AI to group individuals with similar symptoms and behavior suggest that social elements may play a role in enhancing engagement and adherence.[55] However, the direct contribution of these social or behavioral components to sustainable improvements in managing rheumatic diseases is not established. Future studies should prioritize evaluating specific intervention characteristics that maximize efficacy, scalability, and long-term sustainability for individuals with rheumatic diseases in order to clarify how AI may enhance adherence, personalize treatment, or engage patients more effectively.

Finally, as AI tools evolve, sustained collaboration with AI experts and the integration of ethical considerations, including data privacy and accessibility, will be critical for developing robust and equitable AI-supported interventions in rheumatology. This may include research on patient preferences for the use of AI, which can inform ethical design as well as promote improved uptake of these technologies.

In conclusion, this review highlights the emerging role of AI in nonpharmacological interventions for managing chronic rheumatic diseases, with promising applications in education and exercise. Although AI-supported interventions show potential for improving patient engagement, monitoring, and adherence, much of the current research remains in early stages, with limited long-term evidence on clinical effectiveness. Expanding AI applications to address broader aspects of rheumatic care, such as psychological support, nutrition, sleep, and medication adherence, along with clearer definitions and transparent reporting of AI methodologies, will be essential for advancing this field. As AI technology continues to evolve, future studies with diverse populations, rigorous evaluations, and interdisciplinary collaboration will be critical to realizing the full potential of AI in rheumatic disease management.

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 Shah 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. Adelowo O, Mody GM, Tikly M, et al. Rheumatic diseases in Africa. Nat Rev Rheumatol 2021;17(6):363–374.
2. Barber MRW, Drenkard C, Falasinnu T, et al. Global epidemiology of systemic lupus erythematosus. Nat Rev Rheumatol 2021;17(9):515–532.

3. Beydon M, McCoy S, Nguyen Y. Epidemiology of Sjögren syndrome. Nat Rev Rheumatol 2024;20(3):158–169.
4. Finckh A, Gilbert B, Hodkinson B, et al. Global epidemiology of rheumatoid arthritis. Nat Rev Rheumatol 2022;18(10):591–602.
5. Salaffi F, Di Carlo M, Carotti M, et al. The impact of different rheumatic diseases on health-related quality of life: a comparison with a selected sample of healthy individuals using SF-36 questionnaire, EQ-5D and SF-6D utility values. Acta Biomed 2019;89(4):541–557.
6. National Institute of Arthritis and Musculoskeletal and Skin Diseases. Arthritis and Rheumatic Diseases. Updated June 2025. http://niams.nih.gov/health-topics/arthritis-and-rheumatic-diseases
7. Bokhari SFH, Mushtaq A. Psychosocial aspects of rheumatic disease management: addressing mental health and well-being. Cureus 2023;15(11):e49267.
8. Kłak A, Raciborski F, Samel-Kowalik P. Social implications of rheumatic diseases. Reumatologia 2016;54(2):73–78.
9. Hasan O, Jessar M, Ashar M, et al. Systemic sclerosis: clinical manifestations, anesthetic and orthopedic considerations in a patient. Int J Surg Case Rep. 2018;42:24–28.
10. Tanaka H, Okada Y, Nakayamada S, et al. Extracting immunological and clinical heterogeneity across autoimmune rheumatic diseases by cohort-wide immunophenotyping. Ann Rheum Dis 2024;83(2):242–252.
11. Borham L, Hafiz W. Pharmacotherapy in systemic rheumatic diseases. In: Almoallim H, Cheikh M, eds. Skills in Rheumatology. Springer; 2021. https://www.ncbi.nlm.nih.gov/books/NBK585760/
12. Harmsen S, Nabuurs JA, Lehman de Lehnsfeld LF, et al. Mapping the complex everyday challenges and needs of people with rheumatic disease and their surroundings using a multi-actor approach. Musculoskeletal Care 2022;20(4):873–891.
13. Küçükdeveci AA. Nonpharmacological treatment in established rheumatoid arthritis. Best Pract Res Clin Rheumatol 2019;33(5):101482.
14. Parodis I, Girard-Guyonvarc'h C, Arnaud L, et al. EULAR recommendations for the non-pharmacological management of systemic lupus erythematosus and systemic sclerosis. Ann Rheum Dis 2024;83(6):720–729.
15. Moseng T, Vliet Vlieland TPM, Battista S, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis: 2023 update. Ann Rheum Dis 2024;83(6):730–740.
16. Murphy SL, Poole JL, Chen YT, et al. Rehabilitation interventions in systemic sclerosis: a systematic review and future directions. Arthritis Care Res (Hoboken) 2022;74(1):59–69.
17. Effectiveness, safety, and cost-effectiveness of nonpharmacological and nonsurgical therapies for chronic pain. In: National Academies of Sciences Engineering, and Medicine; Health and Medicine Division; Board on Global Health; Board on Health Sciences Policy; Global Forum on Innovation in Health Professional Education; Forum on Neuroscience and Nervous System Disorders. Stroud C, Posey Norris SM, Bain L, eds. The Role of Nonpharmacological Approaches to Pain Management: Proceedings of a Workshop. National Academies Press (US); 2019.
18. Epidemiology of rheumatic diseases. Nature. Updated January 16, 2023. Accessed January 29, 2025. https://www.nature.com/collections/cebbdgddcd
19. Stryker C, Kavlakoglu E. What is Artificial Intelligence (AI)? IBM. Accessed Nov 27, 2024. https://www.ibm.com/topics/artificial-intelligence
20. Brons A, Wang S, Visser B, et al. Machine learning methods to personalize persuasive strategies in mHealth interventions that promote physical activity: scoping review and categorization overview. J Med Internet Res 2024;26:e47774.
21. Inojosa H, Voigt I, Wenk J, et al. Integrating large language models in care, research, and education in multiple sclerosis management. Mult Scler 2024;30(11–12):1392–1401.
22. Bucher A, Blazek ES, Symons CT. How are machine learning and artificial intelligence used in digital behavior change interventions? A scoping review. Mayo Clin Proc Digit Health 2024;2(3):375–404.
23. Sharma M, Savage C, Nair M, et al. Artificial intelligence applications in health care practice: scoping review. J Med Internet Res 2022;24(10):e40238.
24. Cho C-H, Lee T, Lee J-B, et al. Effectiveness of a smartphone app with a wearable activity tracker in preventing the recurrence of mood disorders: prospective case-control study. JMIR Ment Health 2020;7(8):e21283.
25. Rini C, Porter LS, Somers TJ, et al. Automated Internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015;156(5):837–848.
26. Bobrovsky H, Clarno K, Jain D, et al. A decentralized trial evaluating the efficacy of a multimodal pain treatment mobile application for osteoarthritis. Osteoarthr Cartil 2022:30(s1):S371–S372.
27. Piette JD, Newman S, Krein SL, et al. Patient-centered pain care using artificial intelligence and mobile health tools: a randomized comparative effectiveness trial. JAMA Intern Med 2022;182(9):975–983.
28. Mitchell EG, Maimone R, Cassells A, et al. Automated vs. human health coaching: exploring participant and practitioner experiences. Proc ACM Hum Comput Interact 2021;5(CSCW1):99.
29. Harari YN. Nexus: A Brief History of Information Networks from the Stone Age to AI. Random House; 2024.
30. Yelne S, Chaudhary M, Dod K, et al. Harnessing the power of AI: a comprehensive review of its impact and challenges in nursing science and healthcare. Cureus 2023;15(11):e49252.
31. Munn Z, Peters MDJ, Stern C, et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018;18(1):143.
32. Zhao J, Li L, Li J, et al. Application of artificial intelligence in rheumatic disease: a bibliometric analysis. Clin Exp Med 2024;24(1):196.
33. Cipolletta E, Fiorentino MC, Vreju FA, et al. Editorial: artificial intelligence in rheumatology and musculoskeletal diseases. Front Med (Lausanne) 2024;11:1402871.
34. Nelson AE, Arbeeva L. Narrative review of machine learning in rheumatic and musculoskeletal diseases for clinicians and researchers: biases, goals, and future directions. J Rheumatol 2022;49(11):1191–1200.
35. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19–32.
36. Microsoft 365. Microsoft Excel. 2024.
37. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467–473.
38. Schardt C, Adams MB, Owens T, et al. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007;7:16.
39. Innovation VH. Covidence systematic review software. www.covidence.org
40. Buchanan BG, Davis R, Feigenbaum EA. Expert systems: a perspective from computer science. In: Ericsson KA, Charness N, Feltovich PJ, et al, eds. The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press; 2006:87–103.
41. Shalev-Shwartz S, Ben-David S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press; 2014.
42. Chowdhary KR. Natural Language Processing. Fundamentals of Artificial Intelligence. Springer India; 2020:603–649.

43. NetImpact Strategies. Deterministic Process Automation and Artificial Intelligence are not the same. September 28, 2023. Accessed February 04, 2025. https://netimpactstrategies.com/insights/point-of-views/deterministic-process-automation-and-artificial-intelligence-are-not-the-same/
44. Nikzad–Khasmakhi N, Balafar MA, Reza Feizi–Derakhshi M. The state-of-the-art in expert recommendation systems. Eng Appl Artif Intell 2019;82:126–147.
45. Bossen D, Veenhof C, Van Beek KE, et al. Effectiveness of a web-based physical activity intervention in patients with knee and/or hip osteoarthritis: randomized controlled trial. J Med Internet Res 2013; 15(11):e257.
46. Feng Y, Wu Y, Liu H, et al. Effect of the telemedicine-supported multicomponent exercise therapy in patients with knee osteoarthritis: study protocol for a randomized controlled trial. Trials 2023;24(1):729.
47. He J, Ferrigno C, Shakoor N, et al. Responses to gait retraining using pressure-based auditory feedback for medial knee osteoarthritis. Osteoarthr Cartil 2021:29(s1):S81–S82.
48. Pelle T, Bevers K, van der Palen J, et al. Effect of the dr. Bart application on healthcare use and clinical outcomes in people with osteoarthritis of the knee and/or hip in the Netherlands; a randomized controlled trial. Osteoarthr Cartil 2020;28(4):418–427.
49. Ramkumar PN, Haeberle HS, Ramanathan D, et al. Remote patient monitoring using mobile health for total knee arthroplasty: validation of a wearable and machine learning-based surveillance platform. J Arthroplasty 2019;34(10):2253–2259.
50. Thiengwittayaporn S, Wattanapreechanon P, Sakon P, et al. Development of a mobile application to improve exercise accuracy and quality of life in knee osteoarthritis patients: a randomized controlled trial. Arch Orthop Trauma Surg 2023;143(2):729–738.
51. Xu Y, Chen X, Wang L, et al. Efficacy of a computer vision-based system for exercise management in patients with knee osteoarthritis: a study protocol for a randomised controlled pilot trial. BMJ Open 2024;14(11):e077455.
52. You M, Chen X, Liu D, et al. ChatGPT-4 and wearable device assisted Intelligent Exercise Therapy for co-existing Sarcopenia and Osteoarthritis (GAISO): a feasibility study and design for a randomized controlled PROBE non-inferiority trial. J Orthop Surg Res 2024;19(1):635.
53. Blasco JM, Hernández-Guillén D, Pérez-Maletzki J, et al. Design and development of a natural language processing tool to promote adherence to home rehabilitation after major joint replacement surgeries in osteoarthritis. Osteoarthr Cartil 2024:32(6):774.
54. Blasco JM, Díaz-Díaz B, Igual-Camacho C, et al. Effectiveness of using a chatbot to promote adherence to home physiotherapy after total knee replacement, rationale and design of a randomized clinical trial. BMC Musculoskelet Disord 2023;24(1):491.
55. Studenic P, Stamm T, Seidler Y, et al. Rheumabuddy4.0 leading the path to a patient-driven electronic support and monitoring tool. Ann Rheum Dis 2022:81(s1):1801–1802.
56. Scott D, Jansons P, De Ross B, et al. Digital Voice Assistants for delivering exercise, nutrition, and medication information to post menopausal women with osteoporosis: a 24-week pilot randomized controlled trial. Aging Clin Exp Res 2023:35:S154–S155.
57. Tan TC, Roslan NEB, Li JW, et al; Ratnasari. Patient acceptability of symptom screening and patient education using a chatbot for autoimmune inflammatory diseases: survey study. JMIR Form Res 2023;7: e49239.
58. Parra Sanchez AR, Grimberg MG, Hanssen M, et al. Web-based eHealth Clinical Decision Support System as a tool for the treat-to-target management of patients with systemic lupus erythematosus: development and initial usability evaluation. BMJ Health Care Inform 2023;30(1):e100811.
59. Pelle T, Bevers K, van der Palen J, et al. Development and evaluation of a tailored e-self-management intervention (dr. Bart app) for knee and/or hip osteoarthritis: study protocol. BMC Musculoskelet Disord 2019;20(1):398.
60. Scuderi GR, Bourne RB, Noble PC, et al. The new Knee Society Knee Scoring System. Clin Orthop Relat Res 2012;470(1):3–19.
61. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998;28(2):88–96.
62. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation in Industry. CRC Press; 1996:189–194.
63. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001;45(5): 453–461.
64. Couper MP, Tourangeau R, Conrad FG, et al. Evaluating the effectiveness of visual analog scales: a web experiment. Soc Sci Comput Rev 2006;24(2):227–245.
65. Nikiphorou E, Philippou E. Nutrition and its role in prevention and management of rheumatoid arthritis. Autoimmun Rev 2023;22(7): 103333.
66. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97.
67. Wang H, Ho AF, Wiener RC, et al. The association of mobile health applications with self-management behaviors among adults with chronic conditions in the United States. Int J Environ Res Public Health 2021;18(19):10351.
68. Parra Sanchez A, Grimberg MG, Hanssen MMA, et al. Sle-T2t – a digital treat-to-target clinical decision support system for the management of systemic lupus erythematosus patients: development and usability evaluation. Ann Rheum Dis 2022;81(s1):93.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 282–295
DOI 10.1002/acr.25614

American College of Rheumatology
Empowering Rheumatology Professionals

Lived and Care Experiences of Chronic Musculoskeletal Shoulder Pain in Australian Adults: A Qualitative Study

Sonia Ranelli,[1] Joanne E. Jordan,[2] Ilana N. Ackerman,[3] Alison Thorpe,[4] Jennifer G. Persaud,[5] Linda J. Woodhouse,[6] Jason Chua,[7] Ben Horgan,[8] and Andrew M. Briggs[9]

Objective. Australian evidence on lived and care experiences of chronic musculoskeletal shoulder pain (CMSP), irrespective of disorder classification or disease, is limited. However, such evidence is important for person-centered care and informing local service pathways and care guidelines or standards. To address this gap, we explored the lived experiences of adults with CMSP across domains of the International Classification of Function, Disability and Health (ICF) framework, and their care experiences, preferences, and priorities for CMSP.

Methods. This was a qualitative study that applied a phenomenological approach. Purposive sampling was conducted with adults experiencing CMSP. Individual semistructured interviews, informed by ICF domains, explored lived and care experiences/preferences of the participants. Data were analyzed using an inductive approach, by objective.

Results. Twenty adults (50% female) with diverse CMSP conditions/diagnoses, clinical profiles, and age (21–78 years) participated. Five lived experience themes were identified: 1) impact on body functioning; 2) impact on sleep, energy, and drive; 3) impact on mental well-being and evolving sense of self; 4) coping with CMSP; and 5) social support and participation. Four care experience themes were included: 1) care-seeking choices; 2) interactions with health care professionals (HCPs); and 3) values and preferences for components of CMSP care.

Conclusion. Adults with CMSP experience impacts across life stages in multiple domains of functioning (ICF categories) relating to personal and social dimensions. Clinical encounters, particularly interactions with HCPs, influence an individual's confidence and engagement in their CMSP care. Discussion, education, and goal setting through shared decision-making are valued attributes of clinical encounters among people with CMSP.

Musculoskeletal shoulder pain is a common condition, with an estimated median community point prevalence of 16% (range 0.7%–55.2%) across countries.[1] Although the condition is often self-limiting, up to 50% of adults who seek care continue to experience pain and/or functional problems up to two years after symptom onset.[2] A recent Australian study identified that more than two-thirds of younger adults (aged 20–55 years) experiencing shoulder pain reported symptoms exceeding 12 months duration alongside substantial quality of life impairment, high psychological distress, and reduced work participation.[3] CMSP is defined as persistent shoulder pain experienced in musculoskeletal structures for more than three months.

CMSP contributes substantially to the health care burden of musculoskeletal pain in Australia[4] and globally.[5] In Australia, the mean annual cost of health care and domestic support for CMSP is estimated at $7,563 (Australian) per person for patients on orthopedic review waitlists, increasing to $22,378 (Australian) when lost productivity is also considered.[6] National registry data also show that shoulder joint replacement surgery for CMSP in Australia is increasing over time and is projected to exceed

Supported by The Arthritis Foundation Western Australia Incorporated.
[1]Sonia Ranelli, BSc(PT), ClinMPhys(Sports), MSc(Oxon), PhD: Curtin University, Sir Charles Gairdner Hospital, and Perth Children's Hospital, Perth, Western Australia, Australia; [2]Joanne E. Jordan, BA, BSc, MPH, PhD: HealthSense (Australia) Pty Ltd, Melbourne, Deakin University, Geelong, and The Australian Centre for Behavioural Research in Diabetes, Carlton, Victoria, Australia; [3]Ilana N. Ackerman, BPhysio(Hons), PhD: Monash University, Melbourne, Victoria, Australia; [4]Alison Thorpe, BAppSc(PT), MPhil, PhD, FACP: Perth Shoulder Rehabilitation, Perth, Western Australia, Australia; [5]Jennifer G. Persaud, BSc(PT) (Hons), MSc: Sir Charles Gairdner Hospital, Arthritis and Osteoporosis Western Australia, and Hello JP Consulting, Perth, Western Australia, Australia; [6]Linda J. Woodhouse, BSc(PT), BA, MA, PhD: Creighton University, Omaha, Nebraska; [7]Jason Chua, BSc, PhD: Curtin University, Perth, Western Australia, Australia, and Auckland University of Technology, Auckland, New Zealand; [8]Ben Horgan: East Metropolitan Health Service, Perth, Western Australia, Australia; [9]Andrew M. Briggs, BSc(PT)Hons, PhD, FACP: Curtin University, Perth, Western Australia, Australia.

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.25614).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25614.

Address correspondence via email to Andrew M. Briggs, BSc(PT)Hons, PhD, FACP, at A.Briggs@curtin.edu.au.

Submitted for publication December 9, 2024; accepted in revised form July 9, 2025.

SIGNIFICANCE AND INNOVATIONS

- There is limited primary research that evaluates both the lived and care experiences and preferences of people experiencing chronic musculoskeletal shoulder pain (CMSP). Yet, such research is needed to inform person-centred care and service pathways in the local Australian context. Exploring lived experiences through their intersection with the domains of the International Classification of Function, Disability and Health (ICF) Framework enables a more comprehensive understanding of the scope and significance of functioning impacts to inform holistic care. Exploring care experiences and preferences can inform service delivery pathways.
- Diversity in CMSP experiences across gender, age, place of residence, symptom duration, and level of impairment was intentionally sought with maximum heterogeneity sampling.

45,000 procedures annually by 2035 at a cost of over $1.46 billion (Australian).[7]

Although the burden and cost of CMSP have been estimated in Australia, our understanding of the lived experiences of Australian adults with CMSP across multiple domains of functioning coupled with their experiences of, and preferences for, care remain less clear. For example, existing evidence syntheses have not included any primary studies conducted in Australia[8,9] nor have they evaluated preferences for service delivery, yet local evidence acquired through qualitative research is needed to inform local service planning. A rich understanding of both lived experiences and care preferences and priorities are essential for person-centered care and for informing local care planning and delivery at service and system levels.[10,11] This is relevant to Australia given the existing knowledge gap and the structural differences in the Australian health system compared with health systems of other countries. For example, Australia has a dual public and private funding model for health care, and governance responsibilities for hospital and primary care services vary between jurisdictional and national governments.

Existing syntheses of qualitative research highlight that individuals who experience shoulder pain report physical functioning, emotional, and social impacts; hold strong biomechanical beliefs about their shoulder pain; and express fear avoidance.[8,9] Primary studies often sample cohorts with homogenous shoulder disorders, such as rotator cuff-related shoulder pain, yet experiences of CMSP and care preferences are likely to be common across shoulder pathologies (i.e. transdiagnostic)[8,9,12] and aligned with evidence for noncancer pain more generally.[13,14] This suggests that a transdiagnostic conceptualization of CMSP is warranted for informing person-centered care and care pathways at scale. Therefore, studies designed to inform care and care pathways should adopt a sampling frame irrespective of the classification of shoulder disorders. Current evidence also provides limited insight into how lived experiences of CMSP relate to functioning (defined by the ICF) and care priorities and expectations. The ICF provides a valid framework for considering multiple domains of functioning and has been used previously for qualitative explorations of lived experience of functional impacts related to other musculoskeletal conditions, such as arthritis and low back pain.[15–17] Further, although previous quantitative evidence identifies preferences of adults with CMSP related to surgery and primary care interventions,[18,19] there are very limited data from qualitative explorations relating to their values and preferences for care components and therapeutic encounters. This is important because it is well-established that preferences for, and expectations of, care can influence treatment outcomes and care satisfaction among adults with musculoskeletal pain conditions, including CMSP.[20–24] Central to care planning for an individual, and ultimately, for the co-creation of health services, is a knowledge of people's care experiences, expectations, and priorities of care.[25,26] To address these evidence gaps, this study aimed to explore the lived and care experiences and preferences of adults with CMSP using qualitative methods.

METHODS

Design. A cross-sectional design employing qualitative methods, underpinned by a phenomenological approach, was undertaken in Australia. A phenomenological approach is appropriate for exploring individuals' lived experiences and how they make sense of their experiences.[27] Data were collected between February and July of 2022. Findings are reported in alignment with recommendations from the Consolidted criteria for Reporting Qualitative research (COREQ-32) and Guidance for Reporting Involvement of Patients and the Public-version 2 (GRIPP-2) checklists (Supplementary Files 1 and 2). The design considered the exploration of lived experiences and care experiences as separate objectives. The authors' reflexivity and rigor statements are included in Supplementary File 3.

Ethical approval. This study was approved by the Human Research Ethics Committee of Curtin University, Australia (HRE2021-0396). All participants provided informed consent.

Sampling. A purposive sampling frame was used to achieve maximum heterogeneity across demographic criteria, including age, gender, occupation group, geography (metropolitan, regional, rural) and highest level of education. Neither race nor ethnicity were considered in sampling or collected as participant-level characteristics.

In this study, CMSP was defined as a persistent or recurrent pain experience perceived in the bone(s), joint(s), muscle(s), or

related soft/connective tissue(s) of the shoulder complex for longer than three months that was not attributed to an autoimmune inflammatory disease or cancer, and that had necessitated a visit to a health care professional. Eligible participants needed to be aged 18 years or over; community-dwelling and a resident in Australia; currently reporting CMSP in at least one shoulder consistent with the case definition; and able to read and write in English. Participants were excluded if any nonmusculoskeletal shoulder pain condition, such as past/current history of cancer or systemic inflammatory arthritis, was identified through screening. Neck-related shoulder pain was not an exclusion criterion.

Recruitment. Participants were recruited via advertising through community health and leisure centers, arthritis and pain advocacy organizations, a state-level research engagement network, clinical practices and clinical organizations, and social media.

Data collection processes and outcome measures. Individuals who expressed an interest in participating accessed an online survey (Curtin University-licensed Qualtrics, Provo, Utah, USA) that displayed a plain language summary of the research, screened for eligibility, and captured demographic characteristics. Participants who met the inclusion criteria were sampled purposively against demographic criteria to achieve maximal heterogeneity. Participants were contacted via email (SR) and invited to participate in an interview and respond to a short online survey to capture their clinical characteristics using validated patient-reported outcome measures (Table 1).[28–30]

A semistructured interview schedule was developed by the investigators, which included a person with lived experience (BH). The schedule was informed by an earlier Community Conversation with consumers through the Western Australian Health Translation Network,[31] using the World Café method.[32] The initial community conversation identified that lived and care experiences were important to consumers and rationalized the need to explore these dimensions within a research paradigm. The interview schedule was organized into two parts to align with the study objectives: lived experiences and care experiences, expectations and preferences (Supplementary File 4), extending and refining the informal questions used in the Community Conversation. For lived experiences, questions explored the impact of CMSP aligned with key ICF domains (https://icd.who.int/browse/2024–01/icf/en), including: impairments in body structures and body functions; environmental factors; and activities and participation restrictions. For care experiences and preferences, items were informed by contemporary principles of models of care.[11] The schedule was piloted with two people with CMSP and refined before use.

Individual interviews were conducted by SR via video conference (Microsoft Teams) or phone. There were no repeat interviews. All interviews were audio-recorded, transcribed verbatim, and member-checked to allow participants to make any corrections or additional reflections. SR took notes during each interview and met with JEJ and AMB for regular debriefing and reflections.

Data from the first 10 participants were analyzed before further sampling to assess data redundancy and adequate heterogeneity in demographic characteristics. Based on review (SR, AMB, JEJ), it was agreed that further understanding of lived experience of CMSP in younger adults and older men (65+ years) was required, with another 10 interviews conducted. Criteria for ceasing sampling were (1) data redundancy in derived codes and no new information to inform categories and (2) adequate heterogeneity in demographic characteristics.

Data analysis. Sociodemographic and clinical data were analyzed using descriptive statistics (SPSS IBM Corp V29, NY). Analysis of transcripts followed an inductive approach[33] without reference to the ICF. The first 10 transcripts were read several times, with notes made to understand concepts, descriptions, and meanings of phenomena. Descriptive codes were inductively derived by two researchers independently (SR, JEJ). An initial coding framework (Excel sheet) was developed between the two analysts and refined through regular meetings and discussion of their observational and coding notes. Two other researchers

Table 1. Clinical outcome measures*

Construct	Measurement tool	Description
Shoulder pain	Pain NRS	The pain NRS measures average shoulder pain experienced in the past seven days ranging from 0 (no pain) to 10 (worst pain).
Shoulder function	Modified SPADI	The SPADI is a two-domain (pain and function) self-administered tool consisting of five and eight items, respectively, regarding the severity of an individual's pain and the degree of difficulty experienced when performing various activities of daily living requiring upper-extremity use, measured on a 10-item NRS. The means of the two subscales are averaged to produce a total score ranging from 0 (best) to 100 (worst).[28,29]
	*Q*DASH questionnaire	The *Q*DASH is an 11-item tool that measures a person's ability to perform activities of daily living. It produces a score from 0 (no disability) to 100 (greatest disability).[30]

* NRS, numeric rating scale; *Q*DASH, quick version of the Disabilities of the Arm Shoulder and Hand scale; SPADI, Shoulder Pain and Disability Index.

(AMB, INA) independently reviewed and validated the coding framework with a subset of the transcripts (25%), with further refinements made. The coding framework was then used to deductively code the remaining 10 transcripts. Where new concepts were identified, new codes were inductively developed.

Codes were then developed into categories through identifying shared concepts or relationships using an iterative process. Data redundancy was identified after 20 interviews, which was evidenced by no new codes or information to further develop categories. Summary descriptions of categories were developed to identify overarching themes that were discussed and iterated (AMB, JEJ, and SR). The thematic framework was reviewed at a workshop with all investigators and considered against the ICF domains. Participants were provided with a plain language results summary.

RESULTS

Of the 60 people who expressed interest in the study, 37 (61.7%) were screened as eligible and 20 (54%) participated in interviews.

Demographic and clinical characteristics. Demographic and clinical characteristics of the 20 participants are shown in Tables 2 and 3, respectively. There was equal representation between genders with a mean (SD) age of 50.8 (18.7) years (range: 20–78 years). Mean (SD) CMSP duration for the group was 55.5 (55.9) months (range: 4 months to 19 years). Pain duration and the Shoulder Pain and Disability Index and quick version of the Disabilities of the Arm Shoulder and Hand scale outcomes (Table 3) showed substantial variation across the sample (from very high to very low pain and disability), which was consistent with the maximum heterogeneity sampling frame. Most (85.0%) participants resided in metropolitan areas, and the median [interquartile range] decile for relative socioeconomic advantage and disadvantage was high (8 [4]) with a wide range (2–10), based on the Australian Socio-Economic Indexes for Areas (SEIFA) (Table 3).

Table 2. Demographic characteristics of the included sample, N = 20

Characteristic	N (%)
Highest level of education completed	
University post-graduate degree	10 (50)
University undergraduate degree	3 (15)
Trade school or trade college	4 (20)
Secondary/high school	3 (15)
Current work status	
Full time normal duties	6 (30)
Part time normal duties	4 (20)
Full time modified duties	1 (5)
Not working and receiving compensation	1 (5)
Not working (unrelated to shoulder problems)	8 (40)
Usual occupation	
Manager	2 (10)
Professional	7 (35)
Clerical and admin	1 (5)
Sales worker	2 (10)
Retired	7 (35)
Student	1 (5)

Qualitative data. The mean interview length was 29 minutes (range: 19–48 minutes). Figure 1 illustrates a thematic map of findings. Although each research objective was analyzed separately, our findings also signal interdependence between lived and care experiences. Specifically, lived experiences influenced care experiences, particularly care-seeking behaviors, and care experiences influenced lived experiences of CMSP, notably body functioning and coping.

Objective 1: lived experiences. Five key themes were identified, with participants highlighting the impacts of CMSP on body functioning and consequences for valued life activities and sense of self; sleep, energy, and drive; mental well-being and evolving sense of self; coping with CMSP; and social support and participation. These themes were interdependent, as illustrated in Figure 1. The themes are discussed subsequently with salient subthemes italicized. Table 4 provides a summary of themes and all subthemes, and Supplementary File 5 provides additional supporting quotes.

Theme 1.1: impact on body functioning. CMSP was associated with reduced body functioning, particularly physical functioning of the upper limb, which limited participation in valued activities. This theme was prominent among all participants irrespective of age and gender. Participants consistently identified shoulder pain and impairments in range of motion and strength as key barriers to participation. Physical functioning limitations were commonly associated with a sense of frustration and loss that individuals could not participate in activities that were important or meaningful to them. Valued activity restrictions included *independent living and instrumental activities of daily living* (iADLs), including care for children; *work and study*, where for some this required retraining for a new occupation; and *social and leisure participation*, including sport.

Body functioning impact manifested differently across ages. For example, young people (eg, 20–32 years) described "feeling left out" when unable to participate in activities with their peers or family. Older people, such as those in retirement (age 65–76 years), particularly women, were more accepting of participation restrictions: "I also had this feeling that I'm getting older all the time and perhaps it's perfectly reasonable to expect that I don't want to do, or can't do, some of the things I do." (PID01)

Body functioning impacts related to work manifested differently according to occupation type. Participants employed in less physically demanding occupations were better able to perform their duties and retain employment. Some people had to make adjustments, such as better self-care across the working day, whereas for others, CMSP did not restrict their work participation.

Table 3. Clinical characteristics of the included sample, including individual-level (PID01–PID20) and pooled data*

Participants 1–20	Gender	Age (years)	CMSP duration (months)	Shoulder pain diagnosis (self-reported)	*Q*DASH (scale range 0–100)	SPADI pain subscale (scale range 0–100)	SPADI disability subscale (scale range 0–100)	SPADI total score (scale range 0–100)	Geographic area[a]	SEIFA index of relative socioeconomic advantage and disadvantage (decile) in 2021[b]
PID01	Woman	76	60	Osteoarthritis	40.9	42.0	18.8	30.4	Metropolitan (MM1)	6
PID02	Woman	32	108	Unknown	13.6	46.0	26.3	36.2	Metropolitan (MM1)	8
PID03	Man	30	228	Instability	6.8	0.0	3.8	1.9	Metropolitan (MM1)	10
PID04	Woman	71	7	Fracture	20.5	30.0	20.0	25.0	Metropolitan (MM1)	5
PID05	Man	78	60	Osteoarthritis	6.8	12.0	6.3	9.2	Metropolitan (MM1)	10
PID06	Woman	49	12	Frozen shoulder	29.5	44.0	37.5	40.8	Metropolitan (MM1)	9
PID07	Woman	48	38	Rotator cuff-related pain, subacromial bursitis	4.5	16.0	0.0	8.0	Regional (MM2)	5
PID08	Woman	56	6	Frozen shoulder	15.9	32.0	25.0	28.5	Metropolitan (MM1)	9
PID09	Woman	32	24	Unknown	15.9	48.0	48.8	48.4	Metropolitan (MM1)	10
PID10	Man	41	22	Unknown	38.6	56.0	65.0	60.5	Rural (MM3)	5
PID11	Man	53	60	Osteoarthritis	27.3	64.0	35.0	49.5	Metropolitan (MM1)	6
PID12	Man	40	6	Rotator cuff-related pain	0.0	0.0	0.0	0.0	Metropolitan (MM1)	6
PID13	Man	22	12	Subacromial impingement	4.5	28.0	5.0	16.5	Metropolitan (MM1)	3
PID14	Man	62	60	Rotator cuff tear, osteoarthritis	25.0	32.0	36.3	34.2	Regional (MM2)	3
PID15	Woman	77	80	Radiation neuropathy	72.7	56.0	87.5	71.8	Metropolitan (MM1)	2
PID16	Woman	20	120	Unknown	27.3	70.0	45.0	57.5	Metropolitan (MM1)	9
PID17	Woman	64	120	Osteoarthritis	22.7	4.0	7.5	5.8	Metropolitan (MM1)	8
PID18	Man	75	66	Osteoarthritis	13.6	36.0	23.8	29.9	Metropolitan (MM1)	9
PID19	Man	49	4	Rotator cuff-related pain	20.5	48.0	17.5	32.8	Metropolitan (MM1)	10
PID20	Man	40	16	Rotator cuff-related pain	45.5	28.0	45.0	36.5	Metropolitan (MM1)	9
Pooled mean (SD)	–	50.8 (18.7)	55.5 (55.9)	–	22.6 (17.2)	34.6 (20.5)	27.7 (22.9)	31.1 (20.2)	1 (0)[c]	8 (4)[c]
Female mean (SD)	–	52.5 (19.9)	57.5 (46.7)	–	26.4 (19.1)	38.8 (19.1)	31.6 (24.9)	35.2 (20.6)	1 (0)[c]	8 (4)[c]
Male mean (SD)	–	49.0 (18.3)	53.4 (66.3)	–	18.9 (15.2)	30.4 (21.9)	23.8 (21.3)	27.1 (20.2)	1 (0)[c]	7.5 (6)[c]

* CMSP, chronic musculoskeletal shoulder pain; IRSAD, Index of Relative Socio-Economic Advantage and Disadvantage; *Q*DASH: quick version of the Disabilities of the Arm Shoulder and Hand scale; SEIFA, Socio-Economic Indexes for Areas; SPADI, Shoulder Pain and Disability Index.

[a] Geographic area was defined according to the Australian Modified Monash Model (MMM),[34] which defines whether a location is metropolitan, rural, remote or very remote. Areas classified MM 2 to MM 7 are considered regional, rural, or remote.

[b] SEIFA is a product developed by the Australian Bureau of Statistics that ranks areas in Australia according to relative socioeconomic advantage and disadvantage. The IRSAD summarizes information about the economic and social conditions of people and households within an area. This index includes both relative advantage and disadvantage measures. Expressed in deciles (1–10), a low score indicates relatively greater disadvantage and a lack of advantage in general, whereas a high score indicates a relative lack of disadvantage and greater advantage in general [Australian Bureau of Statistics (2021), Socio-Economic Indexes for Areas (SEIFA), Australia, ABS Website, accessed 9 September 2024].[35]

[c] Expressed as median (interquartile range).

Figure 1. Thematic map for both research objectives. Themes for lived experiences were interrelated and a dependency between lived and care experiences emerged.

For people in physically demanding occupations, some had to transition to other employment because their CMSP prevented them continuing in their role.

Theme 1.2: impact on sleep, energy, and drive. Problems with sleep and sleep quality due to CMSP impacted people across genders and age. Sleep impacts included a reduced length and *quality of sleep*, characterized as being unable to fall asleep, and multiple awakenings due to shoulder pain. These sleep impacts manifested as *impaired concentration at work or at school/study tasks*, as well as *fatigue and reduced energy, motivation, and productivity levels*. Impacts were particularly pronounced for younger and middle-aged adults who were employed or studying. They experienced concentration impairments and reduced energy, motivation, and productivity levels. Participants also described temporal changes in their mood, such as feeling "irritable," which were associated with fatigue due to poor or inadequate sleep: "I had a lot of trouble focusing in school. I would rock up to school tired until Year 11…but my mood wasn't that good either. I would be really moody." (PID03)

Theme 1.3: impact on mental well-being and evolving sense of self. Participants consistently cited a change in their self-identity attributed to their CMSP. They reported reduced confidence in their body to perform functional tasks, including sports, coupled with a *fear of reinjury or symptoms exacerbation*. Their loss of confidence and/or fear was represented through a change in their perceived sense of self-worth and general well-being, such as feeling "*damaged*." For some, their loss of confidence also manifested as an emotional toll (eg, feelings of depression and low mood) as they attempted to navigate life with CMSP. However, a few participants expressed the opposite side of the coin, describing a positive sense of well-being. Their perspective seemed to be based on their stage and/or rate of recovery and their care experiences, without negative impacts on mental well-being: "I'm improving every day, and I can't really ask for more than that, I don't think. I'm pretty happy." (PID08)

The nature of mental well-being impacts varied according to how participants experienced participation restrictions, their pain severity, and social context. Participants' expression of their mental health impact varied from those who experienced high distress to others who reflected an acceptance of participation restrictions without distress yet continued to feel some level of *loss*, *sadness*, *and/or frustration*. For example, one young woman described

Table 4. Lived experience theme and subtheme descriptions, including alignment with ICF categories*

Subthemes	Description of subthemes
Theme 1.1: impact on body functioning (ICF categories: body structures; body functions)	
Impact on independent living and iADLs	Daily independent living tasks were more difficult or could not be performed because of CMSP. Functioning limitations were related to pain and impairments in range of movement and strength. Functioning limitations extended to self-care activities, such as washing and dressing, as well as care for children.
Impact on work and study	Work and study tasks, as well as productivity with work or study, were prevented or compromised by range of movement impairments and/or pain.
Social and leisure participation	Lifestyle and leisure activities were restricted or had to be discontinued because of CMSP, often leading to feelings of (or actual) social isolation.
Impact on driving	The ability to drive was restricted, especially related to pain and impairments in range of shoulder movement and strength. Participants described an inability to hold static shoulder postures required for safe driving.
Theme 1.2: impact on sleep, energy and drive (ICF category: body functions)	
Quality of sleep	The quality and length of sleep was negatively impacted because of pain and difficulty in finding comfortable positions. This led to negative impacts on a person's general sense of well-being and increased sensitivity (reduced tolerance) to pain.
Impaired concentration at work (or at school/study)	Reduced sleep duration and quality impaired concentration levels necessary to undertake study or work tasks.
Fatigue and reduced energy, motivation and productivity levels	Ongoing impacts from poor sleep resulted in fatigue that reduced productivity and motivation.
Theme 1.3: impact on mental well-being and evolving sense of self (ICF category: body functions)	
Reduced confidence in body functioning and fear of reinjury and/or symptom exacerbation	Participants cited a decreased confidence in the capacity of their body to perform physical tasks, including sports and leisure activities. Reduced confidence could result in discontinuing tasks or activities due to fear of reinjury or symptoms exacerbation, and manifested as a reduced sense of self-worth, well-being, and for some, an emotional toll.
Fear of disclosure	Participants expressed a fear of disclosing their CMSP in work or social contexts because of how they may be perceived or judged, including a concern about experiencing stigma.
Experiencing social isolation, loss, sadness and/or frustration	When individuals considered there to be no options to adapt activities to accommodate their CMSP and body functioning limitations, they withdrew from relationships and social participation, resulting in low mood, isolation, and loneliness.
Theme 1.4: coping with CMSP (ICF category: activities and participation)	
Learned behavioral responses to CMSP: experiential learning and fear of pain	People with CMSP learned to change their behaviors in response to pain, or the anticipation of pain. For some, they applied experiential learning to perform tasks at home or at work as a strategy to maintain functioning, such as using their unaffected limb for some tasks, performing tasks differently, or supporting their affected limb to function using assistive products. For others, their behaviors changed because of fear of pain, resulting in not using their affected limb at all. This avoidance behavioral strategy was reinforced by experiences of pain exacerbations associated unconscious or unplanned movement. The avoidance behaviors consistently manifested as using their unaffected limb.

(Continued)

Table 4. (*Cont'd*)

Subthemes	Description of subthemes
Acceptance of functioning limitations	Over time, participants reached an acceptance of their functioning limitations, recognizing and accepting that ways of doing tasks and aspects of everyday life were now different.
Social reference comparison and motivation to actively manage their CMSP	Participants compared their functioning limitations and symptoms to others in their social network, recognizing their own situation, in many cases, was less severe. This downward social comparison was used as a coping strategy and motivation to actively manage their condition to maintain quality of life and well-being.
Theme 1.5: social support and participation (ICF category: environmental factors)	
Accessing social support to maintain functioning and overall well-being	Accessing social support, either paid or voluntary, including family and work colleagues, was used differently across participants. Although some participants wished to remain independent, for others who experienced more severe impacts, social support was needed to maintain their functioning. Social networks also provided reassurance that supports were available, if needed.
Impact on social relationships	CMSP impacts could extend beyond personal (intrinsic) factors to also impact external (extrinsic) factors, such as relationships with family. Typically, participants perceived their families carried a burden from their CMSP.
Reengaging in social participation	Participants, particularly those who had become socially isolated, identified the value of social re-engagement to their overall well-being.

* CMSP, chronic musculoskeletal shoulder pain; iADLs, instrumental activities of daily living; ICF, International Classification of Function, Disability and Health.

feelings of mental distress and "hopelessness": "It definitely makes me feel kind of hopeless, more useless. I feel a lot of shame and guilty. I can feel very depressed and a lot of self-anger, because I feel like why can everyone else just go about and do things and function normally, but I can't?" (PID16)

For some, participation restriction in a valued life activity (eg, sports) was associated with a significant mental health impact and shifted their self-identity, for example from being active and socially engaged, to an identify of *social isolation*: "I was sad. I was sad that a lot of my friends were sporty, and I couldn't play basketball with them…I wanted to play with my friends, but I was always left out, standing on the sidelines." (PID03)

For some participants, concerns also extended to a *fear of disclosure* of their CMSP within work or social contexts. Specifically, they feared being judged or experiencing stigma, which also extended to a fear of disclosure in seeking new employment.

Theme 1.4: coping with CMSP. People living with CMSP adopted varied coping strategies or behaviors to continue their usual functioning, either at work or home. This included *learned behavioral responses*, with which individuals tried to improve their functioning, such as pacing, performing a task differently, or using assistive products to support functioning:

"But you learn by trial and error, as well. In the shower, I can't use a proper towel, they're too heavy. I use a camping towel…So that's what I use to dry myself after a shower." (PID15)

However, for others, habits developed as a learned behavioral strategy to avoid exacerbating symptoms. Another coping strategy described was participants arriving at a cognitive and emotional stage of *acceptance of their functioning limitations*, recognizing and accepting that ways of doing tasks and aspects of everyday life had to be different. Some participants were more accepting of their new reality, possibly hopeful of change, whereas others were resigned to their new "normal."

Creating a social reference comparison also served as a coping strategy for some, with individuals comparing their condition to peers whom they perceived to be worse off or experiencing more significant health challenges than themselves (ie, downward social comparison). This served to minimize the perceived negative impacts of their own condition on their sense of well-being. It also served as *motivation to actively manage their CMSP* to improve their quality of life and well-being.

Theme 1.5: social support and participation. Participants identified the importance of *accessing social support to maintain functioning and overall well-being*. Social supports, such as family or colleagues, were identified as important depending on the extent of functioning limitations experienced by the participant and their age. Social supports provided reassurance that assistance was available to maintain their functional ability. However, some older participants expressed a reluctance to seek support for iADLs from their social networks, aiming to remain independent: "Actually, I do not like to call on other people at all, even my friends. I don't like to. I can still vacuum and dust and do the general household things, and I don't want to ever get to the stage where I have to call on people." (PID04)

Table 5. Care experience theme and subtheme descriptions*

Subthemes	Description of subthemes
Key theme 2.1: care-seeking behaviors	
GP as the entry point to care	The GP was the most common first contact HCP for most participants experiencing CMSP, influenced by tradition, prior experience and trust. Some older participants always consulted their GP, while middle-aged participants tended to consult either GP or other HCP based on their previous experience of musculoskeletal pain (shoulder or other area) or recommendations from trusted individuals. Younger participants (20–30- year-olds) often consulted HCPs that their family typically visited. Care pathways from the GP typically followed pharmacological care and onward referral for imaging and nonsurgical interventions, before consideration for specialist referral.
Seeking care from other primary care providers	Where participants were unsatisfied with their care, they sought primary care providers (other than a GP) who were recommended by trusted family and friends.
Key theme 2.2: interactions with HCPs (positive)	
HCPs took time to listen and validate the person's story	Participants described the importance of HCPs acknowledging their concerns and validating the impact of CMSP on their lives. Participants appreciated HCPs being familiar with their case/clinical history and understanding the components of their care journey.
HCPs providing education about their CMSP	Participants valued interactions with HCPs that involved education about their CMSP and their care options, in particular when care options were staged and linked to improving functioning and recovery, and achieving personal goals. Taking an active role in care planning, through shared decision-making, led to a sense of empowerment and motivation to participate in self-care.
Interprofessional communication and coordination of care between HCPs	Participants were complimentary and appreciative of transparent, clear, consistent, and honest communication by their HCPs. They appreciated care delivered by different professions, when this was needed, as well as communication about that care between providers. Participants reflected that this communication enabled HCPs to be aware of the components of care and to better coordinate care.
Amount of time participants are able to spend with HCPs.	Participants relished time with HCPs to better understand their CMSP and receive care, in particular, learning what to do and how, to reduce pain and return to usual function.
Key theme 2.2: interactions with HCPs (negative)	
Lack of engagement and validation	Participants described negative interactions with HCPs as characterized by a lack of engagement and validation, leaving participants feeling confused and uncertain about their care.
Lack of education and goal setting	Participants described a negative interaction with HCPs as one characterized by a one-way or didactic communication style in which care for CMSP was "prescribed" without an opportunity for discussion, education, or shared decision-making to ensure interventions were aligned with goals. Participants perceived that these interactions resulted in care that was not tailored, leading to a delay in recovery and re-engagement in valued life activities.
Lack of interprofessional communication and care coordination between HCPs	Many participants described receiving mixed, contradictory information from HCPs regarding their CSMP, care, and prognosis. This promoted confusion and disillusionment with care and health care system, and for some, a prolonged recovery period.
Inadequate time with HCP	Many participants described time-poor HCPs working within a health care system in which consultation times were short, providing inadequate opportunities for discussion and limiting their understanding about their condition. Short consultation times diminished confidence in their HCP and their ability to make informed care decisions.

(Continued)

Table 5. (*Cont'd*)

Subthemes	Description of subthemes
Key theme 2.3: values and preferences for components of shoulder care	
Preference to avoid pharmacologic interventions if another nonpharmacologic treatment option was likely to provide benefit	Participants' previous experience with, and expectations of, pharmacologic care (eg, side effects of medicines) influenced their treatment preferences. Here, many participants expressed a preference to avoid pharmacologic care if a viable nonpharmacologic option, such as exercise, was available. Exercise was perceived to be safer and was associated with increased confidence in their upper limb and conferring benefits to pain, function, and well-being.
Timely care and face-to-face consultations	Participants' experiences with the COVID-19 restrictions on accessing health care influenced and informed their preferences for care. From these experiences, particularly for residents of Victoria,[a] preferences for timely access to HCPs and for consultations in a face-to-face mode were expressed.

* CMSP, chronic musculoskeletal shoulder pain; HCP, health care professional; GP, general practitioner.
[a] Victoria is a state in Australia where the longest number of days in lockdown were recorded for the COVID-19 pandemic.

Participants also identified that the impacts of living with CMSP extended beyond personal (intrinsic) factors and also affected *social relationships*. Some reflected that their families were burdened by their condition, having to "do the heavy lifting" around the home: "…it's just more stress in the household that you're putting on your partner, when there's enough happening at home. So yeah, again, just conscious that it's putting more stress on her." (PID20)

Objective 2: care experiences. Participants' reflections about their care experiences related to: care-seeking behaviors; positive and negative interactions with health care professionals (HCPs); and values and preferences for components of CMSP care. These themes explored subsequently with salient subthemes italicized and discussed in further detail. Table 5 provides a summary of themes and all subthemes, and Supplementary File 6 provides additional supporting quotes.

Theme 2.1: care-seeking behaviors. Participants described variation in entry points to care. Participants of all ages described trust in and loyalty to their general practitioner (GP), and for most, the *GP was the entry point to care*. Younger and middle-aged participants tended to consult either GPs or other HCPs for their CMSP based on their previous experience(s) of musculoskeletal pain (shoulder or other bodily area) or used recommendations from trusted individuals. Older participants (aged 60–77 years) described more consistently consulting their GP and were influenced by a lack of understanding about what other HCPs could (potentially) offer them: "I think a lot of people my age might say the same thing, that you thought the doctors were gods anyway and whatever they said, you did…We also, rightly or wrongly—and it's wrongly—thought that allied health, like physios and OTs, 'What do they do?' That was a bit of a mystery to us." (PID15)

Among participants who consulted GPs, pharmacologic care and referral for exercise therapy and imaging were typically first-line management. However, for some participants, onward referral for surgical review was initiated following GP consultations and review of imaging.

When participants were less confident or less satisfied with the care they received, they proactively *sought care from other primary care providers* to address their pain and functioning loss. Usually, alternate care pathways were informed by experiences or recommendations from family or friends.

Theme 2.2: interactions with HCPs (positive and negative). Participants expressed both positive and negative care interactions with their HCPs, which included GPs, allied health professionals, and medical specialists. These experiences were centered around HCPs' communication skills, their commitment to education, and shared decision-making that supported care aligned with individuals' goals, time afforded by HCPs, and interprofessional communication and coordination.

Positive experiences were characterized by HCPs with excellent communication skills (verbal and written) and when *HCPs took time to listen to and validate the person's story*, understand their problems, and include them (through shared decision-making) in the care planning and delivery. Participants identified that these characteristics positively influenced their adherence with treatment and optimized their care experiences and outcomes: "I keep going back…I always feel like they're treating me with respect and they're treating me as an intelligent participant in the treatment process, rather than simply a person to be fed information to…." (PID01)

On the other hand, a lack of engagement and validation by HCPs was perceived as a negative interaction in which participants were often left feeling confused and uncertain about their care. Similarly, when participants experienced limited opportunities for discussion and shared decision-making with their HCP, there was potential for misalignment between their recovery expectations and those of their HCP.

Participants valued interactions with HCPs in which they were provided with education about their CMSP. This empowered individuals to actively engage in their care decisions and self-manage. This also promoted confidence in their role in recovery and return to valued activities. Conversely, experiences characterized by a one-way interaction were viewed negatively, resulting in a "*prescription of care*" without the opportunity for discussion, education, and shared decision-making.

Care experiences were often colored by the *amount of time participants were able to spend with their HCPs*. Positive experiences were ones in which participants felt they had adequate time with their HCP to discuss their concerns and understand their condition and care options. When consultation times were inadequate, participants felt less confident and satisfied with their clinical encounter and ill-equipped to make informed care decisions.

Theme 2.3: values and preferences for components of shoulder care. Participants' preferences for components of CMSP care were strongly influenced by their clinical encounters and where they had previously experienced benefit. Many participants, irrespective of age, expressed a preference to *avoid pharmacologic interventions if another nonpharmacologic treatment option was likely to provide benefit*. This preference was largely informed by previous experience with medicines in which little benefit was experienced and side effects outweighed benefits: "Panadol and Nurofen barely touch the edges. I get more side effects from them than I do relief, so I don't bother…" (PID07)

Participants expressed greater confidence in and a preference for exercise therapy. Irrespective of age, exercise was identified as instilling a sense of confidence in using their limb and more sustained benefits for pain, functioning, and a sense of well-being.

DISCUSSION

Adults living with CMSP experience wide-ranging impacts that manifest across four ICF domains (body structures, body functions, activities and participation, and environmental factors), consistent with other musculoskeletal pain research that has interpreted findings in relation to the ICF.[15–17] These impacts resonate with a contemporary understanding of the multidimensional experiences of living with chronic pain,[14,36] as well as condition-specific experiences such as low back pain (LBP) or osteoarthritis.[37,38] Body functioning impacts were strongly identified and interrelated for physical functions (activities of daily living, work/study, leisure, and driving), sleep and vitality, and psychosocial health, highlighting the importance of holistic care for CMSP that considers health and well-being beyond impairments to body structures. The range of coping strategies described (many of which were informed by experiential learning and avoidance behaviors) also point to a need to coach people with CMSP in helpful self-care strategies to optimize their functioning and return to valued life activities. Care experiences were characterized by variability in accessing care and navigating care pathways. Many participants relied on their past experiences, recommendations from social networks, and their GP as their starting points for care.

Our findings on the lived experiences of CMSP support and extend the extant evidence for shoulder pain conditions. Specifically, our findings add to the body of evidence about the multidimensional impacts of shoulder pain disorders, framed through a transdiagnostic perspective, highlighting that experiences and impacts are relevant across diagnoses and classifications. For example, recent qualitative evidence syntheses have identified interrelated physical, psychological, and social impacts of living with various shoulder disorders and nontraumatic shoulder pain.[8,9] However, these reported impacts were not classified using the breadth of ICF domains. Our data also suggest that impact domains are interrelated and, for many impacts, are cascading. For example, body structure impairments limit body functioning and lead to reduced participation and activities; for many, this manifests as mental health impairment, social withdrawal, and isolation. These impacts further compound body functioning impairments, such as sleep disturbance. For some, these impacts were nontrivial, limiting their ability to work and study. These impacts have implications for financial security and the future livelihoods of younger people, including fear of disclosure to employers, consistent with previous investigations.[13,15,39] For many participants, the impact of loss of physical functioning over time influenced their evolving sense of self, such as their self-worth, social connection, and (lack of) confidence in their bodies, including a belief of being "damaged." This belief about structural damage or fragility aligns with previous shoulder pain research[8] and also transcends musculoskeletal pain conditions, commonly expressed by people living with LBP or osteoarthritis.[40–45]

Participants' care experiences reflected attributes of helpful and unhelpful therapeutic encounters that shaped their values and preferences for care. Our findings are consistent with recent evidence,[12] extending it to also feature the importance of validation, consultation time, and interprofessional communication and coordination. Participants strongly preferred consultations that featured validation of their pain experience, a space to share their CMSP-related impacts and goals, opportunities to receive education and support to engage in goal-directed self-care, interprofessional cooperation and collaboration, and adequate consultation time. These therapeutic encounter attributes also align with the values and preferences of people experiencing other musculoskeletal pain conditions that extend over the course of their life[13,46,47] and those that have been identified as independent predictors of satisfaction with GP encounters.[48] Relevant to our findings, a recent evidence synthesis by Klem et al[13] suggested that "helpful" therapeutic encounters could buffer allostatic load associated with sequelae of chronic musculoskeletal pain, whereas "unhelpful" encounters could increase the allostatic load. Similar findings were identified in a review by Eriksen et al[49] who

identified that health care experiences perceived as negative impacted both the physical and mental well-being of patients.

Our findings have important implications for practice and improving service delivery for people living with CMSP. Holistic, person-centered care requires recognition of the dependency between lived experience and care preferences and needs,[11] as illustrated in Figure 1. Our findings also signal that experiences may vary by age, suggesting that stage of life may be an important contextual factor in care planning and delivery. Clinical assessment that is limited to body structure impairment alone is unlikely to elucidate functioning and participation limitations that are meaningful to an individual, resulting in a care pathway that does not align with their expectations or goals. Clinicians should recognize that CMSP is typically associated with multidimensional and interrelated impacts and could actively explore these areas with patient-reported outcome measures (PROMs) as part of clinical consultations. PROMs may be helpful in practice to make sense of a person's impact "loop" or impact "cascade" by evaluating, for example, sleep, mental well-being, coping mechanisms, and social support. Making sense of shoulder pain in a personalized, meaningful, and understandable manner, grounded in an approach that targets emotions, cognitions, and behaviors, is important to improve functioning and self-efficacy, as identified in LBP care.[50] This is important, considering the evidence that patients with shoulder pain are dissatisfied when their expectations and understanding of their experience are misaligned with explanations provided by their health care providers.[8]

Our data also suggest that interprofessional collaboration and consistent communication are valued. Therefore, building transprofessional competencies in chronic pain care will be important to improving care experiences.[51,52] This context, taken together with clear preferences for how care could be optimally delivered, points to an opportunity and targets for optimizing person-centered care for CMSP, such as through the creation of a Clinical Care Standard to drive improvements in health outcomes and care experiences and reduce unwarranted care variation. For example, Australian Clinical Care Standards have been developed for high burden musculoskeletal conditions such as LBP and osteoarthritis.[53,54] Considering wide-ranging impacts and clear preferences for, and variation in, service delivery experiences, our data suggest it may be timely and warranted to consider a Clinical Care Standard for CMSP.

To our knowledge, this is the first primary study conducted in Australia examining both lived and care experiences as well as the care preferences of adults with CMSP. Strengths of this study include its co-creation from inception with people with lived experience, exploration of lived experiences informed by, and interpreted in relation to, ICF domains, and examination of care preferences where current evidence is limited. The purposive sampling frame to achieve balance between genders, wide age distribution, and diversity in clinical profiles enabled more confidence in the transferability of our findings and will likely strengthen the confidence of findings in future evidence syntheses.[9] Nonetheless, our sampling was limited to people largely residing in high socioeconomic areas and underrepresents some groups, such as indigenous Australians, potentially limiting transferability of findings to those communities. Care experiences described by the participants may also be influenced by unique structural characteristics of Australia's health system and Australia's public health response to the COVID-19 pandemic, with extended lockdown periods in the state of Victoria.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the research participants who shared their lived experiences for this study. Open access publishing facilitated by Curtin University, as part of the Wiley - Curtin University agreement via the Council of Australian University Librarians.

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 Briggs 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. Lucas J, van Doorn P, Hegedus E, et al. A systematic review of the global prevalence and incidence of shoulder pain. BMC Musculoskelet Disord 2022;23(1):1073.
2. Linsell L, Dawson J, Zondervan K, et al. Prevalence and incidence of adults consulting for shoulder conditions in UK primary care; patterns of diagnosis and referral. Rheumatology (Oxford) 2006;45(2): 215–221.
3. Ackerman IN, Fotis K, Pearson L, et al. Impaired health-related quality of life, psychological distress, and productivity loss in younger people with persistent shoulder pain: a cross-sectional analysis. Disabil Rehabil 2021;1–10.
4. Australian Institute of Health and Welfare. *The Burden of Musculoskeletal Conditions in Australia: A Detailed Analysis of the Australian Burden of Disease Study* 2011. Welfare AIoH, ed. Canberra: Australian Institute of Health and Welfare; 2017.
5. Gill TK, Mittinty MM, March LM, et al; GBD 2021 Other Musculoskeletal Disorders Collaborators. Global, regional, and national burden of other musculoskeletal disorders, 1990–2020, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5(11):e670–e682.
6. Marks D, Comans T, Bisset L, et al. Shoulder pain cost-of-illness in patients referred for public orthopaedic care in Australia. Aust Health Rev 2019;43(5):540–548.
7. Fisher C, Soh SE, Page RS, et al. Forecasting the future burden of primary total shoulder replacement in Australia. Osteoarthritis Cartilage 2023;31(12):1636–1643.
8. Maxwell C, Robinson K, McCreesh K. Understanding shoulder pain: a qualitative evidence synthesis exploring the patient experience. Phys Ther 2021;101(3):pzaa229.

9. Page MJ, O'Connor DA, Malek M, et al; OMERACT Shoulder Core Set Working Group. Patients' experience of shoulder disorders: a systematic review of qualitative studies for the OMERACT Shoulder Core Domain Set. Rheumatology (Oxford) 2019;58(8):kez046.
10. Belton JL, Slater H, Ravindran TKS, et al. Harnessing people's lived experience to strengthen health systems and support equitable musculoskeletal health care. J Orthop Sports Phys Ther 2023;53(4): 162–171.
11. Speerin R, Needs C, Chua J, et al. Implementing models of care for musculoskeletal conditions in health systems to support value-based care. Best Pract Res Clin Rheumatol 2020;34(5):101548.
12. Maxwell C, Salsberg J, Robinson K, et al. Fostering adherence to evidence-based care in the management of musculoskeletal shoulder pain: a mixed-methods study. Phys Ther 2025;105(3):pzae176.
13. Klem N-R, Slater H, Rowbotham S, et al. Lived and care experiences of young people with chronic musculoskeletal pain and mental health conditions. A systematic review with qualitative evidence synthesis. Pain 2024;166(4):732–754.
14. van Rysewyk S, Blomkvist R, Chuter A, et al. Understanding the lived experience of chronic pain: a systematic review and synthesis of qualitative evidence syntheses. Br J Pain 2023;17(6):592–605.
15. Berkovic D, Ayton D, Briggs AM, et al. "I would be more of a liability than an asset": navigating the workplace as a younger person with arthritis. J Occup Rehabil 2020;30(1):125–134.
16. Esson SA, Côté P, Weaver R, et al. "I stay in bed, sometimes all day." A qualitative study exploring lived experiences of persons with disabling low back pain. J Can Chiropr Assoc 2020;64(1):16–31.
17. Sverker A, Thyberg I, Ostlund G, et al. Participation in work in early rheumatoid arthritis: a qualitative interview study interpreted in terms of the ICF. Disabil Rehabil 2014;36(3):242–249.
18. Smith MV, Gosselin M, Steger-May K, et al. Patient preferences for the treatment of shoulder and proximal biceps disorders are associated with patient age, race, sex, and activity level. Orthop J Sports Med 2018;6(10):2325967118800000.
19. Versloot AHC, Veldwijk J, Ottenheijm RPG, et al. Patient preferences for conservative treatment of shoulder pain: a discrete choice experiment. Family Practice 2024;42(2):cmae050.
20. Bishop MD, Mintken PE, Bialosky JE, et al. Patient expectations of benefit from interventions for neck pain and resulting influence on outcomes. J Orthop Sports Phys Ther 2013;43(7):457–465.
21. Cormier S, Lavigne GL, Choinière M, et al. Expectations predict chronic pain treatment outcomes. Pain 2016;157(2): 329–338.
22. Kongsted A, Vach W, Axø M, et al. Expectation of recovery from low back pain: a longitudinal cohort study investigating patient characteristics related to expectations and the association between expectations and 3-month outcome. Spine 2014; 39(1):81–90.
23. Skatteboe S, Røe C, Fagerland MW, et al. Expectations of pain and functioning in patients with musculoskeletal disorders: a cross-sectional study. BMC Musculoskelet Disord 2017;18(1):48.
24. Lowry V, Desmeules F, Zidarov D, et al. "I wanted to know what was hurting so much": a qualitative study exploring patients' expectations and experiences with primary care management. BMC Musculoskelet Disord 2023;24(1):755.
25. Kravitz RL. Measuring patients' expectations and requests. Ann Intern Med 2001;134(9 Pt 2):881–888.
26. Bowling A, Rowe G, Lambert N, et al. The measurement of patients' expectations for health care: a review and psychometric testing of a measure of patients' expectations. Health Technol Assess 2012; 16(30):i–xii, 1–509.
27. Bazeley P. Qualitative Data Analysis. Practical Strategies. 2nd ed. Sage Publishing; 2020.
28. Breckenridge JD, McAuley JH. Shoulder pain and disability index (SPADI). J Physiother 2011;57(3):197.
29. Williams JW, Holleman DR, Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. J Rheumatol 1995;22(4): 727–732.
30. Beaton DE, Wright JG, Katz JN, Group TUEC; Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005;87(5): 1038–1046.
31. https://wahtn.org
32. https://theworldcafe.com/key-concepts-resources/world-cafe-method/
33. Alase A. The Interpretative Phenomenological Analysis (IPA): a guide to a good qualitative research approach. Int J Educ Lit Stud 2017; 5(2):9–19.
34. https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/2021
35. https://www.health.gov.au/topics/rural-health-workforce/classifications/mmm?language=und
36. Noe-Steinmüller N, Scherbakov D, Zhuravlyova A, et al. Defining suffering in pain: a systematic review on pain-related suffering using natural language processing. Pain 2024;165(7):1434–1449.
37. Bunzli S, Watkins R, Smith A, et al. Lives on hold: a qualitative synthesis exploring the experience of chronic low-back pain. Clin J Pain 2013;29(10):907–916.
38. Wallis JA, Taylor NF, Bunzli S, et al. Experience of living with knee osteoarthritis: a systematic review of qualitative studies. BMJ Open 2019;9(9):e030060.
39. Berkovic D, Ayton D, Briggs AM, et al. "The financial impact is depressing and anxiety inducing": a qualitative exploration of the personal financial toll of arthritis. Arthritis Care Res (Hoboken) 2021;73(5): 671–679.
40. Bunzli S, Smith A, Schütze R, et al. Making sense of low back pain and pain-related fear. J Orthop Sports Phys Ther 2017;47(9): 628–636.
41. Bunzli S, Taylor NF, O'Brien P, et al. Broken machines or active bodies? Part 2. How people talk about osteoarthritis and why clinicians need to change the conversation. J Orthop Sports Phys Ther 2023; 53(6):325–330.
42. Darlow B, Belton J, Brown M, et al. Making sense of osteoarthritis: a narrative review. Osteoarthritis Cartilage 2024;33(1): 17–26.
43. Darlow B, Brown M, Thompson B, et al. Living with osteoarthritis is a balancing act: an exploration of patients' beliefs about knee pain. BMC Rheumatol 2018;2(1):15.
44. Darlow B, Dean S, Perry M, et al. Easy to harm, hard to heal: patient views about the back. Spine 2015;40(11):842–850.
45. Darlow B, Dowell A, Baxter GD, et al. The enduring impact of what clinicians say to people with low back pain. Ann Fam Med 2013;11(6): 527–534.
46. Ames H, Hestevik CH, Briggs AM. Acceptability, values, and preferences of older people for chronic low back pain management; a qualitative evidence synthesis. BMC Geriatr 2024; 24(1):24.
47. Belton J, Birkinshaw H, Pincus T. Patient-centered consultations for persons with musculoskeletal conditions. Chiropr Man Therap 2022; 30(1):53.
48. Basedow M, Hibbert P, Hooper T, et al. Australians with osteoarthritis: satisfaction with health care providers and the perceived helpfulness of treatments and information sources. J Multidiscip Healthc 2016;9: 387–394.
49. Eriksen AA, Fegran L, Fredwall TE, et al. Patients' negative experiences with health care settings brought to light by formal

complaints: a qualitative metasynthesis. J Clin Nurs 2023;32(17–18):5816–5835.

50. Kent P, Haines T, O'Sullivan P, et al; RESTORE trial team. Cognitive functional therapy with or without movement sensor biofeedback versus usual care for chronic, disabling low back pain (RESTORE): a randomised, controlled, three-arm, parallel group, phase 3, clinical trial. Lancet 2023;401(10391):1866–1877.
51. Linton SJ, O'Sullivan PB, Zetterberg HE, et al. The "future" pain clinician: competencies needed to provide psychologically informed care. Scand J Pain 2024;24(1):20240017.
52. Slater H, Jordan JE, O'Sullivan PB, et al. "Listen to me, learn from me": a priority setting partnership for shaping interdisciplinary pain training to strengthen chronic pain care. Pain 2022;163(11):e1145–e1163.
53. Ackerman IN, Doukas F, Buchbinder R, et al. Ensuring a fit-for-purpose resource for consumers, clinicians and health services: the updated Osteoarthritis of the Knee Clinical Care Standard. Med J Aust 2024;221(4):197–200.
54. Maher CG, Archambeau A, Buchbinder R, et al. Introducing Australia's clinical care standard for low back pain. Med J Aust 2023;218(8):354–356.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 296–302
DOI 10.1002/acr.25610

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Did a Non-Medical Biosimilar Switching Policy Cause an Increase in Non-Biologic/Biosimilar Health Care Resource Utilization or Cost in Patients With Inflammatory Arthritis?

HaoHung Dang,[1] Sandra Blitz,[2] Nick Bansback,[3] Michael Law,[4] and Mark Harrison[3]

Objective. This study aimed to evaluate the impact of a series of policies that mandated switching patients with inflammatory arthritis (IA) from an originator biologic to a biosimilar in British Columbia, Canada, on health care resource use and cost.

Methods. The health data of patients with IA were obtained from five linked administrative databases held by Population Data BC from January 2013 to December 2022. Our analysis focused on trends in monthly average use and costs of four care resources: physician services, hospital services, emergency department visits, and concomitant drug use. Using interrupted time series analysis, we evaluated the impact of switching policies targeting (1) infliximab or etanercept and (2) adalimumab on total health care costs, excluding biologic and biosimilar costs.

Results. We included 3,150 patients in the study. Hospital and physician services accounted for the majority of the total care cost for patients with IA. We found no evidence of any increases in physician services, hospital services, emergency department visits, or concomitant drug use after either nonmedical switch policy. We also found no significant change in level and trend in total health care costs for both policies: infliximab or etanercept (level –$40, 95% confidence interval [CI] –$99 to $19; trend $5.42, 95% CI –$0.62 to $11.46) and adalimumab (level –$34, 95% CI –$139 to $70; trend –$8.97, –$17.94 to $0.00).

Conclusion. Nonmedical biosimilar switching policies did not lead to increases in other health care service use and costs.

INTRODUCTION

Biologic drugs have revolutionized the care of people with inflammatory arthritis (IA) because of their ability to reduce symptoms, control disease progression and associated joint damage, and subsequently improve quality of life. However, as a class of drugs, biologics are the biggest expense to public drug plans in many health care systems worldwide.[1] In this context, the arrival of biosimilars, highly similar versions of biologics but at a potentially much lower cost, offered the potential for significant cost savings without a loss of safety and effectiveness. However, the uptake of biosimilars in many jurisdictions, including Canada, following their approval has been slow.[2]

In many health care systems worldwide, nonmedical switching policies have been introduced and have rapidly increased biosimilar uptake.[3–5] In September 2019, British Columbia, Canada, became the first public health care system in North America to announce a nonmedical switching (switching for reasons other than effectiveness, safety, or adherence) policy (known as the Biosimilars Initiative) to increase the uptake of several selected biosimilars.[6] In the first phase, people with IA on originator infliximab or etanercept had to switch to a biosimilar equivalent in

The funders had no role in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

Supported by the Canadian Institutes of Health Research (grant 451221) and the 2020 Canadian Initiative for Outcomes in Rheumatology Care (CIORA) (grant 2020-001). Dr Law's work was supported by the Canada Research Chair in Access to Medicines. Dr Harrison's work was supported by the Michael Smith Foundation for Health Research (Scholar Award 16813).

[1]HaoHung Dang, MSc: University of British Columbia and Centre for Advancing Health Outcomes, Vancouver, BC, Canada; [2]Sandra Blitz, MSc: Centre for Advancing Health Outcomes, Vancouver, BC, Canada; [3]Nick Bansback, PhD, Mark Harrison, PhD: University of British Columbia, Centre for Advancing Health Outcomes, and Arthritis Research Canada, Vancouver, BC, Canada; [4]Michael Law, PhD: University of British Columbia, Vancouver, BC, Canada.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25610).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25610.

Address correspondence via email to Mark Harrison, PhD, at mark.harrison@ubc.ca.

Submitted for publication December 21, 2024; accepted in revised form July 2, 2025.

SIGNIFICANCE & INNOVATIONS

- There has been concern that nonmedical switching biosimilar policies could lead to increased use of other health care resources, such as physician visits, use of concomitant drugs, or hospital and emergency room visits that could offset savings achieved by switching from originator biologics to biosimilars.
- In this study evaluating a nonmedical switching policy for biosimilars in British Columbia, Canada, we found no evidence of an increase in other health care use or costs.
- Any effectiveness of nonmedical switching from biologics to biosimilars in reducing costs for health care systems was not offset by increases in other types of health care resource use in this setting.

2019 to maintain coverage under British Columbia's Public PharmaCare program. A subsequent phase in 2021 required similar switching for people with IA receiving adalimumab.[6] The Biosimilars Initiative policies successfully increased the uptake of biosimilars to between 80% and 90% of biologic prescriptions, including those filled using private insurance plans.[2,7,8]

One concern about nonmedical switching policies that has been raised is that the savings from drug acquisition costs might be offset by increased costs of other health care resource use following a switch. These include additional physician visits to support the switch and any associated problems, concomitant drugs, and hospital visits.[9,10] The objective of this study was to examine whether total nonbiologic or nonbiosimilar health care resource use and associated costs increased at the population level following a nonmedical switching policy in British Columbia, Canada.

MATERIALS AND METHODS

Study setting and design. The Biosimilars Initiative comprised a series of policy changes to increase the uptake of biosimilars for some of the top-spending biologics. This study focuses on two phases of the Biosimilars Initiative nonmedical switching policies: phase one, which covered etanercept and infliximab (May 27, 2019 to November 25, 2019), and the adalimumab phase (April 7, 2021 to October 6, 2021), which coincided with the first approval of an adalimumab biosimilar for PharmaCare.[6] Each policy allowed a six-month transition period for switching patients to biosimilars to maintain provincial coverage. British Columbia's provincial drug coverage is universal but has high deductibles for many households. In addition, approximately 60% of the population hold private insurance coverage, mostly as a benefit of employment.

This study uses descriptive time series to assess trends in components of nonbiologic or nonbiosimilar health care resource use before and after the Biosimilars Initiative policies and an interrupted time series (ITS) analysis to evaluate the impact of these policies on total health care use costs. ITS is a quasi-experimental design commonly used to evaluate the outcomes of an intervention when a randomized trial is not possible. ITS establishes trends in the periods before a policy is introduced and tests whether any statistically significant changes in the level or slope of this trend occur at or after the time of implementation of a policy.

The study was approved by the University of British Columbia Behavioural Research Ethics Board (H20-00252). The reporting of the study follows the Reporting of Studies Conducted Using Observational Routinely-Collected Health Data checklist.[11]

Data sources. We used five British Columbia administrative databases held and linked by Population Data BC, which captured health care services used by the British Columbia population from January 1, 2013 to December 31, 2022. Access to data provided by the data stewards is subject to approval but can be requested for research projects through the data stewards or their designated service providers. The following datasets were used in this study: Consolidation File, Medical Services Plan (MSP), Discharge Abstract Database (DAD), National Ambulatory Care Reporting System (NACRS), and PharmaNet. You can find further information regarding these datasets by visiting the PopData project webpage at https://my.popdata.bc.ca/project_listings/17-030D/collection_approval_dates. All inferences, opinions, and conclusions drawn in this publication are those of the author(s) and do not reflect the opinions or policies of the data steward(s).

Demographic characteristics of the study population were obtained through the Consolidation File, and health care resource use data were obtained through (1) the MSP database, which recorded reimbursement for essential services delivered by fee-for-service physicians to eligible British Columbia residents covered by MSP, classified using *International Classification of Diseases, Ninth Revision* (*ICD-9*) diagnosis codes; (2) the DAD, which captured discharges from acute inpatient facilities (hospitals), classified using *International Classification of Diseases, Tenth Revision, Canada* (*ICD-10-CA*) diagnosis codes; (3) NACRS, which recorded ambulatory care data including emergency department (ED) visits, day surgery, and medical and surgical day clinics within hospitals, the community, and private clinics; (4) PharmaNet, which recorded drug dispensation data from community pharmacies and hospital outpatient pharmacies. Unique ED visits were captured from a combination of NACRS, MSP, and DAD using a standard approach.[12]

Study population. People with IA were identified from 2013 to 2022 using a published identification algorithm.[13] Briefly, to be classified as IA, an individual needed to have at least one

hospital record or three physician visits (at least one specialist visit) with an *ICD-9* or *ICD-10-CA* diagnosis code for an IA condition within a two-year period based on published algorithms (see Supplementary Table S1). Treatment with infliximab, etanercept, or adalimumab was determined using PharmaNet data. People with at least two prescriptions of adalimumab, infliximab, or etanercept covering at least 84 nonoverlapping days in a 12-month period before the first day of the policy switch period for a specific drug were included in the study cohort. We excluded people with fewer than two prescriptions of the biologic of interest or less than 84 days of coverage. Members of the cohort were observed from their index date of diagnosis or January 1, 2015, whichever was earliest, until their death, last recorded date in administrative databases, or December 31, 2022, whichever was earliest. An individual is included in the analysis (in the numerator and/or the denominator) for any given month in which they had an existing prescription for one of the drugs of interest.

Outcomes. The outcomes of interest were the major components of health care resource use and the total cost of this health care resource use, excluding the cost of biologics or biosimilars. The four major components of health care resource use were visits to hospitals, physicians, and EDs and concomitant drug use. The mean cost of each component of health care resource use was costed from the public health care payer perspective and summed to represent the total health care resource use cost per person.

Hospital and ED visits were categorized as "related" or "unrelated to switching" based on the *ICD-9* and *ICD-10* codes associated with IA (see Supplementary Table S2). Inpatient hospitalizations were costed using a case-mix approach, which multiplies the cost of a standard hospital stay by the resource intensity weight corresponding to the case-mix group designated by the Canadian Institute for Health Information. ED visits were costed by multiplying the number of unique visits by an ED facility cost. Physician visits were classified into two categories of "specialist" (rheumatologist) and "other" (eg, internal medicine or general practitioner) visits. A maximum of one visit per day per physician was counted for each patient, that is, two visits to the same physician on the same day were counted as one. We used the reimbursement amount from the MSP database for fee-for-service physicians during a visit. Concomitant drug use was measured as the number of prescribed drug days for concomitant drugs (ie, one drug day is one prescribed day for one concomitant drug) and costed by combining the amount paid by PharmaCare and the out-of-pocket costs to individuals. Concomitant drugs included conventional synthetic disease-modifying antirheumatic drugs, nonsteroidal anti-inflammatory drugs, and oral glucocorticoids (see Supplementary Table S3).

Patient-level data were aggregated into monthly data by splitting any prolonged hospitalization or prescriptions (>1 day) evenly over the service use period. Mean care costs for individuals were calculated by dividing the total monthly cost of the health care resource for the cohort by the yearly cohort size. All costs were converted to 2022 Canadian dollars.

Exposure. Exposure was defined as being on a specific drug affected by the Biosimilars Initiative at the time of the nonmedical switch. Patients were included in a yearly cohort if they were on one of the three anti–tumor necrosis factor drugs in that year (adalimumab, etanercept, or infliximab). The policy date for each drug was set at the start of the nonmedical switching period for a given drug: May 27, 2019, for etanercept and infliximab and April 7, 2021, for adalimumab. The analysis covered the period from January 1, 2015 to December 31, 2022, and provided at least one year of data before the approval of the first biosimilar under PharmaCare in British Columbia (infliximab in February 2016).

Analytical approach. Trends of monthly health care resource use over the study period were summarized using descriptive time series. The ITS analysis sought to determine whether there were any significant changes in the level and trend of the total cost of health care resource use at the time of either of the policy start dates; the counterfactual assumption is that in the absence of nonmedical switching policies, health care resource use cost would continue along the preswitch level and trend. The six-month transition periods for each policy were excluded from the ITS analysis, and linear segmented regression models were fitted to the remaining data. The ITS model generated four coefficients (two for each policy) estimating the changes in the level and trend of total health care resource use cost, excluding biologic or biosimilar acquisition cost, before and after the phase one (infliximab and etanercept) and adalimumab phase nonmedical switching in the Biosimilars Initiative. The presence of nonstationarity and autocorrelation between monthly data points was assessed by performing the Durbin–Watson test, autocorrelation function, and partial autocorrelation function. The final model was fitted using a generalized least-squares approach. The likelihood-ratio test was used to assess moving average and autoregressive correlation terms to make the appropriate adjustments to the autocorrelation structures in our model. We conducted the ITS analysis separately for rheumatoid arthritis (RA) and ankylosing spondylitis (AS) groups as a sensitivity analysis. All analyses were conducted using SAS version 9.4 and R version 4.0.5.

RESULTS

We included 3,150 patients with IA in our study cohort. The annual cohort size increased steadily from 1,606 in 2015 to 1,960 in 2022. The cohort was 50% women, and the mean age was 49.8 (SD 15.3) years (see Supplementary Table S4). In April 2019, before phase one of the Biosimilars Initiative nonmedical

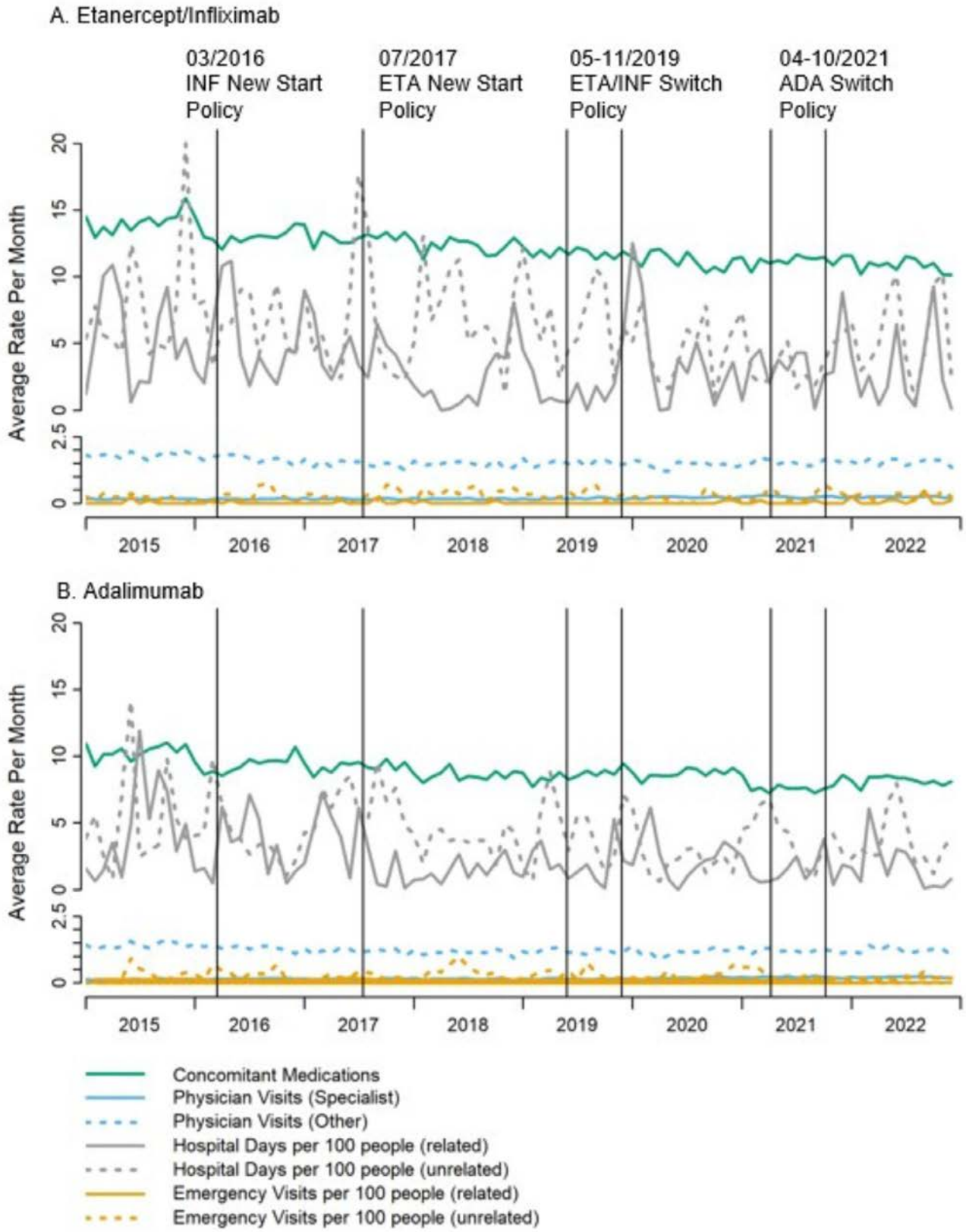


Figure 1. Health care resource use rate per month from 2015 to 2022 for individuals treated with (A) ETA or INF and (B) ADA. ADA, adalimumab; ETA, etanercept; INF, infliximab.

switching policy, 46% of the IA cohort were on infliximab or etanercept, and 25% of these patients were on a biosimilar. The proportion on a biosimilar rose to 86% by January 2020. In March 2021 (before phase two of the Biosimilars Initiative nonmedical switching), 65% of the cohort was on adalimumab (all originator), and by November 2021, 81% had switched to the biosimilar. There were no noticeable changes in the resource use of patients with IA for physician services, hospital services (hospitalization days and ED visits), and concomitant drug days after switching to biosimilar versions of infliximab or etanercept in phase one of the Biosimilars Initiative (Figure 1A) or following the adalimumab phase (Figure 1B).

The cost of hospital visits and physician services represented most of the health care costs (Figure 2). The cost of hospitalizations decreased during the study period, and so, before 2018, the cost of hospitalizations accounted for the majority of total health care resource use costs, but after 2018, physician services became the costliest component of health care resource use.

The results of the ITS on total health care resource use costs are shown in Figure 3 and Supplementary Table S5. In the preintervention period, the intercept for the mean monthly care cost per person with IA was $350, and there was a statistically significant downward trend in total health care resource use cost per month (–$0.84, 95% confidence interval [CI] –$1.65 to –$0.03). After phase one of the Biosimilars Initiative nonmedical switching policy for infliximab and etanercept, there was no statistically significant change in the level (–$40.19, 95% CI –$99.41 to $19.04) or trend ($5.42, 95% CI –$0.62 to $11.46) of total health care resource use costs per person with IA per month. Similarly, for the adalimumab phase of the Biosimilars Initiative nonmedical switching, there was no statistically significant change in the level (–$34.33, 95% CI –$138.86 to $70.19) or trend (–$8.97, 95% CI –$17.94 to 0.00) in total health care resource use cost per person with IA per month. Sensitivity analyses in the RA and AS subgroups found consistent results of no increases in total health care resource use following either nonmedical switching period (see Supplementary Table S6 and Supplementary Figures S1 and S2).

DISCUSSION

Nonmedical switching policies have increased the uptake of biosimilars, offering payers the opportunity to make substantial savings in drug acquisition costs and/or leverage additional competition to secure savings on biologic medications. Importantly, this study presents evidence that any savings will not be offset by increases in spending on other types of health care resource use. This suggests that the magnitude of savings anticipated by payers, for example, the $732 million realized savings in pharmaceutical costs for the British Columbia government in the first five years of the Biosimilars Initiative, are unlikely to be reduced through spending on other health services.[14] These findings provide strong evidence from the first nonmedical switching policy in North America, suggesting other jurisdictions ought not to be concerned with offsets reducing the potential savings of similar nonmedical switching policies.

This study builds upon and supplants a previous series of early initial rapid monitoring studies in phase one of the Biosimilars Initiative, which primarily reported no signals of sustained changes in other types of health care resource use switching cohorts compared with historical cohorts.[7,8] We found no changes in phase two with the introduction of an adalimumab biosimilar switching policy, which is consistent with reports from Denmark following a national nonmedical switch from adalimumab originator to biosimilar.[15] Furthermore, we found no evidence of the temporal, anticipated increase in visits to

Figure 2. Health care resource use costs from 2015 to 2022. ADA, adalimumab; ETA, etanercept; INF, infliximab.

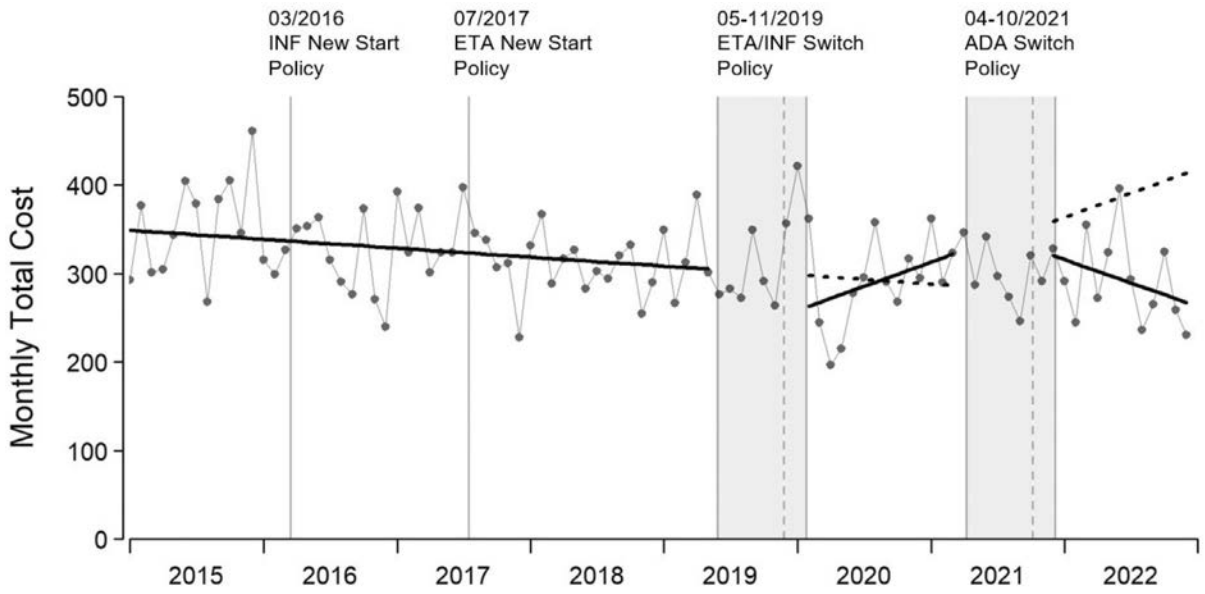


Figure 3. Interrupted time series analysis on total monthly health care resource use care costs. The solid line represents the interrupted time series model estimates, and the dashed lines represent the counterfactual fit (ie, what would have been expected without the policy change). ADA, adalimumab; ETA, etanercept; INF, infliximab.

rheumatologists for monitoring patients following nonmedical switching in our data series.[7,8]

There are some limitations to this analysis. We are unable to quantify the magnitude of cost savings to the British Columbia government after considering all costs of health care resource use and drug acquisition because the negotiated unit prices for biologics and biosimilars are confidential. However, because we excluded the acquisition costs of biosimilars and biologics from this analysis, our finding of no significant change in other health care resource use costs means that organizations with access to the prices paid for biologics and the discount for biosimilars can confidently generate estimates of cost savings.

Despite the limitation, this was a comprehensive population-based study including all patients with IA covered by the British Columbia health care system affected by nonmedical switching under the Biosimilars Initiative. Secondly, our analytical approach includes descriptive trend analyses coupled with robust quasi-experimental statistical methods. Third, the patient cohort used in this was observed over longer study periods than have previously been conducted and included an examination of health care resource use following nonmedical switching of adalimumab for the first time.

In summary, our findings suggest that cost savings from switching to cheaper biosimilars will not be offset by the costs of increased use of other health care services. These findings should provide confidence for other jurisdictions in North America and worldwide that nonmedical switching policies can be effective for cost containment and do not lead to consequences for people switched from originators to biosimilars resulting in increased contacts with the health care system and additional costs to payers.

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 Harrison 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. Makurvet FD. Biologics vs. small molecules: drug costs and patient access. Med Drug Discov 2021;9:100075.
2. McClean AR, Cheng L, Bansback N, et al. Uptake and spending on biosimilar infliximab and etanercept after new start and switching policies in Canada: an interrupted time series analysis. Arthritis Care Res (Hoboken) 2023;75(9):2011–2021.
3. Glintborg B, Sørensen J, Hetland ML. Does a mandatory non-medical switch from originator to biosimilar infliximab lead to increased use of outpatient healthcare resources? A register-based study in patients with inflammatory arthritis. RMD Open 2018;4(2):e000710.
4. Glintborg B, Ibsen R, Bilbo REQ, et al. Does a mandatory non-medical switch from originator to biosimilar etanercept lead to increase in healthcare use and costs? A Danish register-based study of patients with inflammatory arthritis. RMD Open 2019;5(2):e001016.
5. Humphreys SZ. Real-world evidence of a successful biosimilar adoption program. Future Oncol 2022;18(16):1997–2006.
6. British Columbia Ministry of Health. Biosimilars Initiative for Health Professionals. British Columbia Ministry of Health; 2025. Accessed

September 22, 2025. https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/pharmacare/programs/biosimilars-initiative-health-professionals

7. Fisher A, Kim JD, Carney G, et al. Rapid monitoring of health services use following a policy to switch patients from originator to biosimilar etanercept-a cohort study in British Columbia. BMC Rheumatol 2022;6(1):5.
8. Fisher A, Kim JD, Dormuth CR. Mandatory nonmedical switching from originator to biosimilar infliximab in patients with inflammatory arthritis and psoriasis in British Columbia: a cohort study. CMAJ Open 2022; 10(1):E109–E118.
9. Liu Y, Yang M, Garg V, et al. Economic impact of non-medical switching from originator biologics to biosimilars: a systematic literature review. Adv Ther 2019;36(8):1851–1877.
10. Hillhouse E, Mathurin K, Bibeau J, et al. The economic impact of originator-to-biosimilar non-medical switching in the real-world setting: a systematic literature review. Adv Ther 2022;39(1):455–487.
11. Benchimol EI, Smeeth L, Guttmann A, et al; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015;12(10):e1001885.
12. Peterson S, Wickham M, Lavergne R, et al. Methods to Comprehensively Identify Emergency Department Visits Using Administrative Data in British Columbia. Centre for Health Services and Policy Research; 2021.
13. Widdifield J, Bernatsky S, Bombardier C, et al. Rheumatoid arthritis surveillance in Ontario: monitoring the burden, quality of care and patient outcomes through linkage of administrative health data. Healthc Q 2015;18(3):7–10.
14. British Columbia Ministry of Health. B.C.'s Biosimilar Initiative Saves Money, Reinvests to Expand Drug Coverage. British Columbia Ministry of Health; 2024. Accessed September 22, 2025. https://archive.news.gov.bc.ca/releases/news_releases_2020-2024/2024HLTH0081-000927.htm
15. Nabi H, Ibsen R, Ibsen M, et al. Counting the costs: a nationwide study on healthcare use following an adalimumab biosimilar switch in >1300 inflammatory arthritis patients. Ther Adv Musculoskelet Dis 2024;16:1759720X241289391.

Arthritis Care & Research
Vol. 78, No. 2, February 2026, pp 303–304

American College of Rheumatology
Empowering Rheumatology Professionals

LETTER

DOI 10.1002/acr.25633

Statistical considerations regarding the association of higher levels of high-sensitivity C-reactive protein with future development of psoriatic arthritis in psoriasis: comment on the article by Eder et al

To the Editor:

We read the study[1] by Eder et al with great interest. The authors aimed to assess whether high-sensitivity C-reactive protein (hsCRP) levels predict the development of psoriatic arthritis (PsA) in patients with psoriasis. Their multivariable analysis suggested that elevated systemic inflammation, measured by hsCRP, is associated with future PsA development. However, we wish to raise two statistical concerns regarding this study:

1. Sample size and model stability: Current epidemiologic guidelines recommend approximately 10 events per variable (EPV) to ensure model stability.[2–4] With 57 PsA cases, the analysis should include no more than six variables. The inclusion of 11 variables in the multivariable analysis (Table 2 from Eder et al[1]) exceeds this recommendation, potentially compromising model reliability. For a robust analysis of 11 variables, a minimum of 110 PsA cases (10 EPV × 11 variables) would be required.
2. Variable selection for multivariable analysis: Multivariable models should include only variables demonstrating statistical significance in univariate analyses or clinically relevant covariates. Including nonsignificant variables, as in Table 2 from Eder et al,[1] risks model overfitting and may yield inaccurate risk estimates.

Additionally, as laboratory specialists, we recommend incorporating other markers of inflammation (eg, neutrophil or leukocyte counts) into the analysis. Relying solely on hsCRP may overlook the contribution of traditional blood parameters and overstate its predictive role in PsA development.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25633.

Juan Cao, MD
Department of Clinical Laboratory
First Affiliated Hospital of Gannan Medical University
Ganzhou, Jiangxi, China
Jinlin Liu, MD, PhD
liujinlinhz@163.com
Department of Clinical Laboratory
South China Hospital, Medical School, Shenzhen University
Shenzhen, Guangdong, China

1. Eder L, Li X, Chandran V, et al. Association of higher levels of high-sensitivity C-reactive protein with future development of psoriatic arthritis in psoriasis: a prospective cohort study. Arthritis Care Res (Hoboken) 2025;77(11):1288–1292.
2. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
3. Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38(7):1276–1296.
4. Pate A, Riley RD, Collins GS, et al. Minimum sample size for developing a multivariable prediction model using multinomial logistic regression. Stat Methods Med Res 2023;32(3):555–571.

DOI 10.1002/acr.25634

Reply

To the Editor:

Cao and Liu raise concerns about the analyses we conducted examining the relationship between high-sensitivity C-reactive protein (CRP) measurements and the development of psoriatic arthritis (PsA) in patients with psoriasis,[1] citing guidelines regarding the number of covariates appropriate for the development of predictive models.[2] Our objective was not to develop a predictive model but, rather, to adjust for potential confounding variables. In the first column of Table 2 from our previous study,[1] we report the estimated hazard ratio from a Cox model including only CRP. The rationale for reporting results from an adjusted Cox model (see the second column of Table 2) was to address concerns about potential confounding. Individuals contributing to our analyses were required to be PsA-free at the time of the CRP assessment. Accordingly, the CRP assessment time was a left-truncation time, marking the point in which individuals began to be at a risk period for PsA development in our sample. The adjusted analyses also relaxed the assumption of independent left truncation and showed that the relation between CRP and time to PsA in the univariate analyses was not unduly influenced by confounding variables because the effects were comparable and the statistical significance remained strong.

The second statistical concern raised by Cao and Liu relates to variable selection for multivariable analysis. They argue that only variables demonstrating statistical significance in univariate analyses, or those that are clinically relevant, should be included. We considered all variables included in the multivariable model to be clinically relevant, as all have been associated with increased PsA risk.[3,4] In addition, there are multiple reasons why clinically

important covariates may fail to achieve statistical significance in univariate analyses, including sample size and the covariate distribution. Moreover, influential confounders may not show significant univariate associations with the outcome yet still exert an important confounding effect. Therefore, we did not perform variable selection based on statistical significance. Instead, we specified an agreed upon list of potential confounders and adjusted for them. This approach aligns with best practices for addressing confounding, in which the objective is to strengthen causal inference,[5] and we suggest the reported confidence intervals be examined as reflections of the uncertainty of the estimates. Although we acknowledge some effects of confounders were imprecise, our primary aim was to assess the robustness of the hazard ratio for the CRP effect.

Their final point pertained to incorporating other markers of inflammation (eg, neutrophil or leukocyte counts) into the analysis. As we only had access to stored serum samples from this cohort, cell counts could not be analyzed as suggested. We intentionally focused on high-sensitivity CRP, a routinely available and clinically used biomarker, and did not extend the study to include other more advanced protein or metabolite biomarkers that would require access to research laboratories.

ChatGPT-4o was used to check the spelling and grammar of this letter.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25634.

The study was supported in part by a New Investigator Grant from the Physician Services Incorporated (PSI) Foundation. LE is supported by Canada Research Chair in Inflammatory Rheumatic Diseases (Tier 2). VC is supported by a salary award from the Department of Medicine, University of Toronto. The Gladman-Krembil Psoriatic Disease program at Toronto Western Hospital is supported by the Krembil Foundation.

Lihi Eder, MD, PhD
lihi.eder@wchospital.ca
Women's College Research Institute, Women's College Hospital and Department of Medicine and Institute of Medical Science, University of Toronto
Toronto, Ontario, Canada
Xianwei Li, PhD
Richard J. Cook, PhD
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, Canada
Vinod Chandran, MBBS, MD, DM, PhD
Department of Medicine and Institute of Medical Science and Department of Laboratory Medicine and Pathobiology
University of Toronto
Schroeder Arthritis Institute, University Health Network and Department of Medicine, Toronto Western Hospital
Toronto, Ontario, Canada
Cheryl F. Rosen, MD
Department of Medicine and Institute of Medical Science
University of Toronto
and Department of Medicine, Toronto Western Hospital
Toronto, Ontario, Canada
Dafna D. Gladman, MD
Department of Medicine and Institute of Medical Science
University of Toronto
Schroeder Arthritis Institute, University Health Network and Department of Medicine, Toronto Western Hospital
Toronto, Ontario, Canada

1. Eder L, Li X, Chandran V, et al. Association of higher levels of high-sensitivity C-reactive protein with future development of psoriatic arthritis in psoriasis: a prospective cohort study. Arthritis Care Res (Hoboken) 2025;77(11):1288–1292.
2. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ 2020;368:m441.
3. Eder L, Lee KA, Chandran V, et al. Derivation of a Multivariable Psoriatic Arthritis Risk Estimation Tool (PRESTO): a step towards prevention. Arthritis Rheumatol 2025;77(6):686–695.
4. Perez-Chada LM, Haberman RH, Chandran V, et al. Consensus terminology for preclinical phases of psoriatic arthritis for use in research studies: results from a Delphi consensus study. Nat Rev Rheumatol 2021;17(4):238–243.
5. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10(1):37–48.