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. 4 April 2026

Arthritis Care & Research

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

Cover design: Sandra Pulmano

♾ This journal is printed on acid-free paper.

Arthritis Care & Research

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

VOLUME 78 • APRIL 2026 • NO. 4

Cover image: The image on the cover (Halstead et al; page 547–556) shows the proportion of bones with reported bone marrow lesions in a group with midfoot pain at baseline and change of bone marrow lesions at 12-week follow-up for orthosis and control groups.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 431–438
DOI 10.1002/acr.25644

Effectiveness and Safety of Baricitinib for Juvenile Idiopathic Arthritis–Associated Uveitis or Chronic Anterior Antinuclear Antibody–Positive Uveitis

Athimalaipet V. Ramanan,[1] Catherine M. Guly,[2] Gabriele Simonini,[3] Stuart Y. Keller,[4] Priyanka Sen,[4] Thorsten Holzkaemper,[4] Joana Araújo,[4] and Pierre Quartier[5]

Objective. To evaluate the efficacy and safety of baricitinib in pediatric patients with active juvenile idiopathic arthritis–associated uveitis (JIA-U) or chronic anterior antinuclear antibody–positive uveitis, who had an inadequate response to methotrexate (MTX) or biologic disease-modifying antirheumatic drugs (bDMARDs).

Methods. JUVE-BRIGHT was an open-label, active-controlled phase-3 multicenter trial that used a novel design, including 1:1 randomization to an active reference arm. The primary efficacy endpoint was the proportion of responders at week 24 (W24), defined according to the Standardization of Uveitis Nomenclature (SUN) criteria as a two-step decrease in the level of inflammation (anterior chamber cells) or decrease to zero through W24 in the most severely affected eye at baseline. Study success was based on a prespecified Bayesian success rule: the study was deemed successful if there was >80% posterior probability that the baricitinib SUN criteria response rate at W24 was at least 57%.

Results. This study enrolled 30 pediatric patients. The study primary endpoint was not met. In the baricitinib group, 36.8% of MTX–inadequate responder (MTX-IR) and bDMARD-IR patients and 20% of MTX-IR patients achieved a two-step decrease in SUN criteria at W24. Eight patients (33.3%) achieved a response at W24, resulting in 1.03% posterior probability of a response rate of >57%. Safety data were consistent with the established safety profile in other baricitinib indications in pediatric and adult patients.

Conclusion. Although the primary endpoint was not met, the data provide important information on baricitinib for the treatment of children with JIA-U refractory to both MTX and bDMARDs. Baricitinib safety profile in this study was consistent with previous studies in children and adults with other diseases.

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is the most common rheumatic disease of childhood, with JIA-associated uveitis (JIA-U) being the most common extra-articular manifestation of the disease. In patients with JIA, the prevalence of uveitis ranges from 11.6%[1] to 30%.[2] Uveitis is an inflammation of the uveal components of the eye, which includes the iris, choroid, and ciliary body.[3] Despite the well-documented link between JIA and uveitis, the reason for uveal inflammation is not well understood. It is likely that the pathophysiology of JIA-U involves both genetic and environmental elements.[3] A number of risk factors for JIA-U have been identified, which include sex, JIA category, age at onset, and anterior antinuclear antibody (ANA) and HLA-B27 positivity.[2,4–6] Children with chronic ANA-positive uveitis without systemic features (rash, fever, generalized lymph node enlargement, hepatomegaly, splenomegaly, or serositis) have a similar clinical course and treatment response to children with a diagnosis of JIA-U.[7]

The first line of treatments for JIA-U is topical glucocorticoids. Following failure of adequate control of inflammation of topical

ClinicalTrials.gov identifier: NCT04088409.
Supported by Eli Lilly and Company.
[1]Athimalaipet V. Ramanan, FMedSci: Bristol Royal Hospital for Children and Translational Health Sciences, Bristol, United Kingdom; [2]Catherine M. Guly, MB, ChB: Bristol Eye Hospital, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, United Kingdom; [3]Gabriele Simonini, MD: ERN ReConnet Center, Meyer Children's Hospital IRCCS and University of Florence, Florence, Italy; [4]Stuart Y. Keller, BSc, Priyanka Sen, MSc, Thorsten Holzkaemper, MD, Dr med, Joana Araújo, PhD: Eli Lilly and Company, Indianapolis, Indiana; [5]Pierre Quartier, MD: RAISE Reference Centre for Rare Diseases, Necker-Enfants Malades Hospital, Assistance Publique-Hopitaux de Paris and Université Paris-cité, Paris, France.
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.25644).
Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25644.
Address correspondence via email to Athimalaipet V. Ramanan, FMedSci, at a.ramanan@bristol.ac.uk.
Submitted for publication January 23, 2025; accepted in revised form August 21, 2025.

SIGNIFICANCE & INNOVATIONS

- This study investigates baricitinib treatment for children with juvenile idiopathic arthritis–associated uveitis or chronic antinuclear antibody–positive uveitis who are inadequate responders (IRs) to methotrexate (MTX) or IRs to both MTX and biologic disease-modifying antirheumatic drugs.
- The study design used is the first adaptive design trial with a comparator arm in a rare disease in children, which may be beneficial for future pediatric studies.
- This study presents the use of a novel Bayesian design to overcome recruitment challenges for a rare pediatric disease.

glucocorticoids or in children with severe uveitis or ocular complications, the use of disease-modifying antirheumatic drugs (DMARDs) are employed. Methotrexate (MTX) is the first-line systemic therapy.[6] However, patients with moderate to severe JIA-U can be refractory to MTX, and for these patients, biologic DMARDs (bDMARDs) are initiated.[8–10] Adalimumab, an anti–tumor necrosis factor alpha (anti-TNFα) monoclonal antibody, is the only licensed biologic therapy for JIA-U. Although there is a high rate of primary treatment response with adalimumab in JIA-U[11,12] and uveitis control with adalimumab is greater than infliximab and etanercept in the clinical setting,[13] discontinuation of treatment has been reported due to reactivation of uveitis and development of anti-adalimumab antibodies.[14,15] Furthermore data from the ORCHIDEA registry reported 40% of JIA-U patients receiving adalimumab did not achieve clinical remission during a two-year treatment period.[16] Therefore, in patients that do not respond to MTX and anti-TNF drugs, new therapeutic approaches are required. Recently, an anti–interleukin-6 receptor, tocilizumab, has been investigated as a therapeutic with MTX in anti-TNF refractory JIA-U patients. Although the prespecified criterion was not met, 33% had a two-step improvement in the level of inflammation (anterior chamber [AC] cells) at week 12, and a further 14% had a one-step improvement at week 24.[17] This highlights the lack of alternative bDMARDs to treat JIA-U and the need to investigate different mechanisms of action.

Case reports and small series of patients have suggested a benefit of JAK inhibitors, including baricitinib, in adult patients with JIA-U.[18–21] Baricitinib has demonstrated clinical safety and efficacy in adult patients with moderate to severe active rheumatoid arthritis (RA) in phase 3 studies,[22,23] as well as in pediatric patients with polyarticular course JIA.[24] As the etiology and pathogenesis of JIA, JIA-U, and ANA-positive uveitis are still poorly understood but share several immunologic abnormalities identified in RA, it is hypothesized that baricitinib may inhibit JAK enzymatic activity associated with JIA-U and ANA-positive uveitis pathogenesis, thus providing a novel therapeutic approach.[7] Therefore, the aim of this study is to evaluate the efficacy and safety of oral baricitinib administered once daily to pediatric patients with JIA-U or chronic ANA-positive uveitis without systemic features who had an inadequate response to topical steroids, as well as MTX or bDMARDs. Additionally, this study presents a novel approach to the study design of pediatric trials, including the addition of a reference arm and the use of Bayesian analysis.

MATERIALS AND METHODS

Study design. JUVE-BRIGHT was an open-label, active-controlled phase 3 trial conducted at 23 centers, including children from 2 to <18 years old with JIA-U or chronic ANA-positive uveitis without systemic features (ie, rash, fever, generalized lymph node enlargement, hepatomegaly, splenomegaly, or serositis) who had an inadequate response to topical steroids, as well as MTX or bDMARDs. The study protocol has been previously published including study design and eligibility criteria.[7] Briefly, patients were eligible to enroll in the study only if they met several criteria at screening and at baseline, including the following:

1. Were at least 2 years and less than 18 years of age; full date of birth was collected except in countries in which it is not allowed.
2. Had a diagnosis of JIA-U or chronic ANA-positive uveitis without systemic features.
3. Had active anterior uveitis, defined as cellular infiltrate in the AC of Standardization of Uveitis Nomenclature (SUN) criteria grade ≥1+ at visit 1 (screening) and visit 2 (potential randomization), despite prior treatment with adequate doses of topical steroid therapy and MTX.
4. Had an inadequate response or intolerance to MTX (minimum dose of 10 mg/m^2/week, with a maximum dose of 25 mg/m^2/week). Patients considered to have an inadequate response must have received MTX for at least 12 weeks before an inadequate response could have been determined and must have been on a stable dose for at least 4 weeks before screening if continuing MTX therapy during the study.
5. Had been receiving topical glucocorticoid eye drops at a stable dose for at least two weeks before screening (maximum of four drops/day per eye at screening).

Individuals who were taking or who might have been taking JIA or JIA-U medications or treatments interfering with the ability to assess the safety and efficacy of baricitinib were excluded from the study.

At baseline, patient race was self-reported from a fixed set of categories. This study includes adalimumab in the form of an active-control reference arm. The aim was to recruit at least 20 patients who had an inadequate response or intolerance to MTX who would be randomized in a 1:1 ratio to open-label baricitinib or adalimumab (reference arm). Approximately an additional

20 patients who were MTX–inadequate responders (MTX-IRs) or bDMARD-IRs would be offered baricitinib.

The study design included a treatment period up to 24 weeks and an open-label extension period up to 260 weeks. The data presented in this manuscript are for the treatment period up to 24 weeks only. A total of 30 patients were enrolled in the study (Supplemental Figure 1). One patient discontinued before receiving treatment. Of the 29 patients who received treatment, the first 10 patients that were MTX-IRs and bDMARD naive were randomly assigned to baricitinib or adalimumab. The remaining 19 patients were MTX-IRs and bDMARD-IRs and were assigned to baricitinib treatment. Patients received oral baricitinib once daily at age-based doses (9 to <18 years old: 4 mg; 2 to <9 years old: 2 mg). The reference arm of patients, who were MTX-IRs and bDMARDs-naive received adalimumab by subcutaneous injection every two weeks (20 or 40 mg based on weight). Trial registration number is NCT04088409.

Outcomes. The primary efficacy endpoint was the proportion of responders at week 24, defined according to the SUN criteria as a two-step decrease in the level of inflammation (AC cells) or decrease to zero through week 24 in the eye most severely affected at baseline. A Bayesian analysis was performed for the primary endpoint, based on prespecified success criteria (the posterior probability of the treatment response rate exceeding 57% is at least 80%). Response rates of a two-step decrease in SUN criteria in baricitinib-treated patients (who were MTX-IRs and bDMARD-IRs), baricitinib-treated patients who were MTX-IRs, and adalimumab-treated patients who were MTX-IRs are reported. Furthermore, two-step and one-step improvements in SUN criteria are reported over time. Secondary outcomes included change in SUN grade of cells in the AC, Pediatric American College of Rheumatology (PedACR) 30 criteria, PedACR50, PedACR70, PedACR90, and PedACR100 at week 24.[25] Safety evaluations of baricitinib in pediatric patients with JIA-U or ANA-positive uveitis included treatment-emergent adverse events (TEAEs). TEAEs were summarized using descriptive statistics.

Statistical analysis. An open-label Bayesian design was used for this study to maximize the benefit/risk balance for trial participants and to overcome the statistical inference challenge for small sample-size studies. A Bayesian analysis determined if the study would be considered a positive study or a negative study. The study was considered positive if the posterior probability was at least 80% that the baricitinib treatment response would be >57%; otherwise, the study was considered negative.

A Bayesian posterior probability was calculated based on the number of observed responders who completed or had the opportunity to complete 24 weeks of the study. The details of the calculation of the posterior probability are as follows:

Let X_i denote the response of the ith patient, where $X_i|\theta \sim Bern(\theta)$ and $\theta \sim Bet(1,1)$ Then,

$$Y^{stage3} = \sum_{I=1}^{30} X_i$$

The Bayesian posterior probability for study success at the end of study (Stage 3) was calculated as

$$pr\left(\theta > 0.57 \mid \gamma\boxed{stage3}\right) = \frac{Pr\left(\gamma\boxed{stage3} \mid \theta > 0.57\right) . Pr\left(\theta > 0.57\right)}{Pr\left(\gamma\boxed{stage3}\right)}$$

The primary endpoint was imputed using modified NonResponder Imputation (mNRI). Of note, usage of an mNRI method can be justified based on the composite strategy (ICH E9[R1]) for handling intercurrent events.

In a composite strategy with NRI, a patient is defined as a responder only if (1) they meet the clinical requirements for response at the predefined time, and (2) they remain on the assigned study treatment; failing either criterion by definition makes them a nonresponder. For Study JAHW, this definition was modified such that, if either one of (1) or (2) criterion failed, the patient data would be reviewed by the Clinical Endpoints Committee (CEC) for further responder evaluation. The primary efficacy outcome variable, that is, the proportion of responders at week 24, in which response is defined according to the SUN criteria as a two-step decrease in the level of inflammation (AC cells) or decrease to zero in the eye most severely affected at baseline, was imputed using mNRI.

Patients were thus considered nonresponders for the mNRI analysis if they did not meet the clinical response criteria or entirely missed the visit at the analysis time due to lack of efficacy. If the primary efficacy outcome variable data were missing for reasons other than this, available preceding data from patients with missing data were reviewed by an independent and external CEC, as outlined in the CEC charter. The CEC ensured that patients with missing data were evaluated in a consistent manner. The CEC considered the complete patient history and reasons for missing data in order to determine whether responder or nonresponder status was appropriate on a case-by-case basis.

Continuous data are summarized in terms of the observed mean, SD, minimum, maximum, and median. For categorical variables, NRI was used with 95% confidence intervals.

Ethics. This trial was conducted in accordance with the ethical principles of the Declaration of Helsinki and Good Clinical Practice guidelines. The design was approved by the European Medicines Agency's Paediatric Committee (PDCO; EudraCT 2019-000119-10) and by various country-level regulatory agencies and ethical review boards, including the CEIm del Hospital Universitario y Politécnico la Fe (Valencia, Spain), Comitato Etico San Martino (Genoa, Italy), Comitato Etico Milan Area 2 (Milan, Italy), CPP Nord Ouest IV—Lille (Lille, France), Landesamt für Gesundheit und Soziales (Berlin, Germany), and North West–

Table 1. Baseline demographics of the baricitinib- and adalimumab-treated cohorts*

Treatment	Baricitinib (N = 24)	Adalimumab (N = 5)
Age, median (min, max), y	12.0 (6, 17)	7.0 (4, 9)
Weight, median (min, max), kg	47.2 (17.3, 92.1)	22.3 (17.0, 40.9)
Height, median (min, max), cm	155.9 (107.0, 179.1)	123.5 (103.5, 138.1)
BMI, median (min, max)	19.3 (14.6, 33.5)	16.0 (14.6, 21.5)
Female, n[a] (%)	14.0 (58.3)	5 (100.0)
Male, n[a] (%)	10.0 (41.7)	0 (0)
Race, n[a] (%)		
White	20.0 (100.0)	4.0 (80.0)
Black or African American	0 (0)	1.0 (20.0)
Missing	4.0 (0)	0 (0)
Ethnicity, n[a] (%)		
Hispanic or Latino	2.0 (8.7)	0 (0)
Not Hispanic or Latino	12.0 (52.2)	4.0 (80.0)
Not reported	9.0 (39.1)	1.0 (20.0)
Missing	1.0 (0)	0 (0)

* BMI, body mass index; N, number of patients in population.
[a] Number of patients with non-missing data, used as denominator.

Liverpool Central Research Ethics Committee (Liverpool, United Kingdom). All parents or legal guardians and patients (as appropriate) provided written informed consent and assent, respectively. Patient confidentiality will be maintained in accordance with the data privacy statement in the informed consent form.

Data availability statement. Lilly provides access to all individual participant data collected during the trial, after anonymization, with the exception of pharmacokinetic or genetic data. Data are available to request six months after the indication studied has been approved in the US and European Union and after primary publication acceptance, whichever is later. No expiration date of data requests is currently set once data are made available. Access is provided after a proposal has been approved by an independent review committee identified for this purpose and after receipt of a signed data sharing agreement. Data and documents, including the study protocol, statistical analysis plan, clinical study report, blank or annotated case report forms, will be provided in a secure data sharing environment. For details on submitting a request, see the instructions provided at www.vivli.org.

RESULTS

Recruitment to the trial proved challenging, particularly the subpopulation of patients who were MTX-IRs to be randomized to baricitinib or adalimumab (reference arm), because only 10 of the planned 20 patients were recruited. Of the 29 patients

Table 2. Baseline disease characteristics of the baricitinib-treated cohort*

Disease characteristic	Baricitinib (N = 24)		
	Total (N = 24)	MTX-IRs and bDMARD-IRs (N = 19)	MTX-IRs (N = 5)
JIA-U and ANA+, n (%)	8 (33.3)	6 (31.6)	2 (40.0)
JIA-U and ANA−, n (%)	12 (50.0)	10 (52.6)	2 (40.0)
ANA+ without JIA, n (%)	4 (16.7)	3 (15.8)	1 (20.0)
Time since uveitis diagnosis (y), median (min, max)	3.9 (0, 15)	5.7 (0, 15)	2.8 (0, 5)
Number of active joints, mean (SD)	0.5 (1.34)	0.7 (1.5)	0
Number of joints with limited range of motion, mean (SD)	0.3 (0.83)	0.3 (0.84)	0.4 (0.89)
AC SUN grading in the most severely affected eye, SUN grade, n (%)			
0.5+	1 (4.2)	0	1 (20)
1+	7 (29.2)	5 (26.3)	2 (40)
2+	7 (29.2)	5 (26.3)	2 (40)
3+	4 (16.7)	4 (21.1)	0
4+	5 (20.8)	5 (26.3)	0
Physician's global assessment,[a] mean (SD)	3.9 (3.23)	4.2 (3.07)	3.0 (3.98)
Parent's global assessment,[b] mean (SD)	31.8 (27.77)	34.0 (28.27)	24.0 (27.30)
CHAQ Physical Function (score), mean (SD)	0.3 (0.50)	0.4 (0.53)	0.2 (0.34)
ESR, mean (SD)	18.6 (27.85)	21.5 (30.14)	6.0 (8.04)
JADAS-27, mean (SD)	8.3 (4.38)	8.7 (4.59)	6.5 (3.21)
Prior MTX use, n (%)	24 (100.0)	19 (100.0)	5 (100.0)
Prior bDMARD use, n (%)	19 (79.2)	19 (100.0)	0
Number of prior bDMARD used, n (%)			
1	13 (54.2)	13 (68.4)	0
2	3 (12.5)	3 (15.8)	0
>2	3 (12.5)	3 (15.8)	0

* AC, anterior chamber; ANA, anterior antinuclear antibody; bDMARD, biologic disease-modifying antirheumatic drug; CHAQ, Childhood Health Assessment Questionnaire; ESR, erythrocyte sedimentation rate; IR, inadequate responder; JADAS, Juvenile Arthritis Disease Activity Score; JIA-U, juvenile idiopathic arthritis–associated uveitis; MTX, methotrexate; N, number of patients in population; n, number of patients in the specified category; SD, standard deviation; SUN, Standardization of Uveitis Nomenclature.
[a] Physician's Global Assessment of Disease Activity.
[b] Parent's Global Assessment of Well-Being.

Table 3. 2-Step improvement in SUN grade of cells in the AC response rate at week 24 for the baricitinib-treated and adalimumab-treated cohorts*

	Baricitinib			
	Total N = 24	MTX-IRs and bDMARD-IRs N = 19	MTX-IRs N = 5	Adalimumab (MTX-IRs) N = 5
Response at wk 24, n (%)	8 (33.0)	7 (36.8)	1 (20.0)	4 (80.0)
Posterior probability of response rate >57%	1.03%	N/A	N/A	N/A

* AC, anterior chamber; bDMARD, biologic disease-modifying antirheumatic drug; IR, inadequate responder; MTX, methotrexate; N, number of patients in population; n, number of patients in the specified category; N/A, not applicable; SUN, Standardization of Uveitis Nomenclature.

receiving treatment, 24 patients (82.8%) received baricitinib and 5 patients (17.2%) received adalimumab (Table 1 and Supplemental Table 1). Of the baricitinib-treated patients, 19 (79.2%) were MTX-IRs and bDMARD-IRs, and 5 (20.8%) were MTX-IRs alone (Table 2). At baseline, for patients receiving baricitinib (n = 24), the median age was 12 years, 58.3% were female, and the vast majority were White (Table 1).

Of the 19 patients that were bDMARD-IRs, 13 patients (68.4%) had prior experience with one bDMARD, 3 patients (15.8%) had prior experience with two bDMARDs, and 3 patients (15.8%) had prior experience with more than two bDMARDs (Table 2). Regarding disease characteristics, 20 patients had JIA-U and 4 had ANA-positive chronic anterior uveitis without JAI-U (Table 2). Of the 24 patients, 15 treated with baricitinib completed 24 weeks of the study.

For the primary outcome of a two-step decrease in SUN criteria in the baricitinib group, 33% of patients achieved a response at week 24, resulting in 1.03% posterior probability of a response rate of >57% (Table 3); therefore, the study success criteria were not met. For baricitinib-treated patients who were MTX-IRs and bDMARD-IRs, 36.8% of patients were responders. Additionally of the patients that were MTX-IRs, 80% of adalimumab-treated patients responded and 20% of baricitinib-treated patients were responders.

With regard to a step improvement in the SUN criteria, for baricitinib-treated patients, there was a greater proportion of patients who achieved a one-step improvement compared to a two-step improvement across all time points (Table 4). Baricitinib displayed a rapid onset of effect with 70.8% of patients achieving a one-step improvement at week 4. Although this high effect was not sustained, 45.8% of patients still achieved a one-step improvement at week 24.

For the secondary outcomes, at week 24, baricitinib-treated patients had a mean change in SUN grade of cells in the AC of −1.47 (SD: 1.55) in the most severely affected eye and −0.27 (SD: 0.80) in the less severely affected eye (Supplemental Table 2). With regard to PedACR, at week 24, 29.2% of patients treated with baricitinib achieved PedACR30, and 20.8%, 20.8%, 8.3%, and 4.2% of patients achieved PedACR50, PedACR70,

Table 4. Response rate of 1-step and 2-step improvement in SUN grade of cells in the AC for the baricitinib-treated cohort*

	Baricitinib (N = 24)	
Week	SUN criteria 1-Step improvement, n (%)	SUN criteria 2-Step improvement, n (%)
4	17 (70.8)	9 (37.5)
8	14 (58.3)	7 (29.2)
12	16 (66.7)	6 (25.0)
16	15 (62.5)	10 (41.7)
20	14 (58.3)	11 (45.8)
24	11 (45.8)	8 (33.3)

* AC, anterior chamber; N, number of patients in population; n, number of patients in the specified category; SUN, Standardization of Uveitis Nomenclature.

Table 5. Treatment-emergent adverse events occurring in at least 2 patients for the 24-week safety population for the baricitinib-treated cohort*

Treatment-emergent adverse events	Baricitinib (N = 24), n (%)
Infections and infestations	11 (45.8)
COVID-19	5 (20.8)
Nasopharyngitis	4 (16.7)
Upper respiratory tract infection	2 (8.3)
Gastrointestinal disorders	9 (37.5)
Nausea	5 (20.8)
Vomiting	4 (16.7)
Abdominal pain upper	3 (12.5)
Diarrhea	2 (8.3)
Respiratory, thoracic, and mediastinal disorders	7 (29.2)
Cough	4 (16.7)
Oropharyngeal pain	4 (16.7)
Eye disorders	6 (25.0)
Uveitis	2 (8.3)
General disorders and administration site conditions	5 (20.8)
Pyrexia	5 (20.8)
Musculoskeletal and connective tissue disorders	5 (20.8)
Arthralgia	2 (8.3)
Nervous system disorders	5 (20.8)
Headache	4 (16.7)
Injury, poisoning, and procedural complications	4 (16.7)
Ligament sprain	2 (8.3)
Investigations	4 (16.7)
Skin and subcutaneous tissue disorders	3 (12.5)
Acne	2 (8.3)

* COVID-19, coronavirusdisease 2019; N, number of patients in population; n, number of patients in the specified category.

PedACR90, and PedACR100, respectively (Supplemental Table 3). Adalimumab-treated patients' secondary outcomes are reported in Supplemental Tables 4–7.

The overall safety data are shown in Table 5. The most frequent TEAE reported with baricitinib treatment was infections and infestations in 11 patients (45.8%). Two serious adverse events were reported during the trial (uveitis and intentional overdose of paracetamol), and one discontinuation from the study was due to an adverse event (worsening of uveitis). In the baricitinib group, 83.3% of patients reported at least one TEAE of which 41.7% were mild, 29.2% were moderate, and 12.5% were severe. Detailed lists of TEAEs are shown in Supplemental Tables 7–8.

DISCUSSION

Although there is a pressing need to provide a novel therapeutic approach for children with JIA-U refractory to both MTX and bDMARDs, head-to-head or noninferiority studies are not possible for pediatric rheumatic diseases due to the rarity of diseases and the associated sample sizes that would be required. Historical comparisons with other placebo-controlled trials do not address the key issue that access to care and therapeutic paradigms change every couple of years. This trial presents a novel approach through the inclusion of a reference arm of a standard of care biologic agent, which helps address issues important to patients and caregivers and clinicians.

In line with this, MTX-IR patients and their parents showed a low interest in study enrollment due to the known efficacy of adalimumab and its known availability and recommendation as the first-line therapy for MTX-IR patients. Participation in the trial was also declined due to ethical concerns regarding random assignment to the experimental drug over the standard of care.

Additionally, daily dosing of baricitinib may also have been a limiting factor given that some parents prefer fortnightly injections over daily treatment. Enrollment completion in the MTX-IR cohort was thus infeasible, and only 10 patients were randomized to the MTX-IR cohort. The remaining 19 patients assigned to the baricitinib treatment were selected from the bDMARD-IR cohort, as these patients were directly assigned to baricitinib without randomization. The Bayesian design use thus proved to be judicious to overcome statistical inference challenges linked to small sample sizes.

Although a limitation of the study is the small numbers due to recruitment challenge, especially for the randomized arm, strategized data analysis approaches based on the mNRI method and novel Bayesian design allowed us to overcome challenges related to clinical trials with small sample sizes. The mNRI method, in which missing data due to lack of efficacy are imputed as nonresponders, was used. The CEC reviewed available preceding data to determine responder or nonresponder status on a case-by-case basis. Any missing primary efficacy data for reasons other than lack of efficacy were thus imputed as nonresponders. However, there may be biases due to the nonnull risk of overestimating nonresponse rates, as patients who are lost to follow-up are automatically considered nonresponders. This could lead to an underestimation of the treatment's efficacy. Lastly, imputation of missing data in such a small sample size has the potential to impact results, as each patient's data carry more weight in the overall analysis.

To decrease the impact of low patient recruitment in pediatric studies, a novel Bayesian design was used in order to overcome statistical inference challenges related to the small sample sizes. In the context of rare diseases—in which the number of eligible patients is inherently limited—a Bayesian approach enables more efficient use of data that can realistically be collected. Specifically, it allows for the incorporation of prior information (eg, expert opinion, historical data from related trials) to supplement limited trial data and to draw probabilistic conclusions about treatment effects even when sample sizes are small. This stands in contrast to traditional frequentist approaches, which often require larger sample sizes to achieve adequate power. Thus, although Bayesian analysis does not solve recruitment limitations, it provides a statistical framework that makes small-sample trials feasible and interpretable when recruitment is constrained.[26]

Overall, in this study, most patients treated with baricitinib primarily had complex uveitis that failed to be controlled by one or several targeted biologic treatments including adalimumab (Supplemental Table 2). Of note, the adalimumab and baricitinib MTX-IR cohorts had a slightly less severe distribution of AC cell count (Table 2 and Supplemental Table 9) but, overall, similar to the SYCAMORE study population (Table 1).[11] On the other hand, the MTX-IR and bDMARD-IR cohort had a considerably more severe AC cell distribution, with a higher percentage of patients having 3+ and 4+ grades.

Additionally, in patients who are difficult to treat, the severity of disease and previous failure of response to therapy would make it harder to achieve disease control. In line with this, in a previous case study of patients with JIA-U who were refractory to prior conventional DMARDs and bDMARDs and subsequently treated with JAK inhibitors, patients showed improvement of uveitis, which was defined as a reduction of intraocular inflammation according to SUN criteria. However, the mean follow-up period for this case study was 7 months (range 4–13 months)[19]; therefore, a 24 week follow-up for the primary outcome presented here may have underestimated the time for onset of efficacy in difficult-to-treat MTX-IR and bDMARD-IR patients.

Although the study success criteria were not met, it is noted that 36.8% of baricitinib-treated patients that were MTX-IRs and bDMARD-IRs achieved the primary efficacy outcome. Moreover, of the patients that were MTX-IRs, 80% of adalimumab-treated patients responded and 20% of baricitinib-treated patients were responders.

It is not possible to attribute the different outcomes to baseline characteristics in terms of AC cell distribution. A potential explanation to the greater response rate observed in patients

who had failed both MTX and bDMARD than in patients who had failed MTX alone is that these patients may exhibit distinct disease mechanisms. In line with this, for bio-naive patients with mild–moderate disease, TNF seems to be the primary driver of inflammation, which would explain the good response rate obtained with adalimumab. On the other hand, for TNF refractory patients with moderate–severe disease, baricitinib does seem to work, coherently with APTITUDE results.[17] Another plausible factor to explain these results could be the influence of study design and patient selection criteria on the observed outcomes. The inclusion of biologic-IR patients was driven by an unmet clinical need, whereas patients who had failed MTX monotherapy and were subsequently assigned to baricitinib in an open-label setting may have had a preference for adalimumab. This does not exclude the possibility of bias within an already limited sample size.

Although the primary endpoint was not met, benefits were observed with baricitinib treatment, given that across all time points, a greater proportion of patients achieved a one-step improvement. These results thus support previous findings obtained in small series of patients with JAK inhibitors being beneficial for the treatment of patients with JIA-U.[18–21]

In this JUVE-BRIGHT study, the safety profile of baricitinib was found to be consistent with the established baricitinib indications in pediatric and adult patients.[22–24] Baricitinib also offers the advantage of being an oral drug, which may increase compliance in juvenile patients. Given that most patients had complex uveitis with a lack of response to one or more targeted treatments including adalimumab and only a third of patients responded to baricitinib, it may offer clinicians with additional information on treatment approaches for children with JIA-U who have failed both MTX and adalimumab. Although this study did not meet the primary outcome, the data provide additional information for the treatment of children with JIA-U refractory to both MTX and bDMARDs. Overall, results from JUVE-BRIGHT demonstrated that small portion of patients with JIA-U or chronic anterior ANA-positive uveitis that were MTX-IRs and bDMARD-IRs did achieve clinical response, which may be informative for clinicians to consider when treating for children who fail MTX and adalimumab.

ACKNOWLEDGMENTS

Writing assistance was provided by Caroline Murphy, PhD, and Anissa Jhumka, PhD, employees of Eli Lilly and Company. The authors would like to acknowledge Samuel Ogwu, PhD, and Hitendra Pandey, MS, employees of Eli Lilly and Company, for statistical review and support. The authors thank Rona Wang, MD, MBA, and Ana Paula Accioly, MD, MBA, employees of Eli Lilly and Company, for their review and scientific contribution to the manuscript.

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, Mr Ramanan 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

Eli Lilly and Company had sole responsibility of study design, collection of data, analysis and interpretation of the data. Eli Lilly and Company did involve clinical experts as key consultants in the design of the study as well as interpretation of the data. Eli Lilly and Company was responsible for drafting the manuscript and submitting it for publication.

REFERENCES

1. Angeles-Han ST, Pelajo CF, Vogler LB, et al; CARRA Registry Investigators. Risk markers of juvenile idiopathic arthritis-associated uveitis in the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Registry. J Rheumatol 2013;40(12):2088–2096.
2. Moradi A, Amin RM, Thorne JE. The role of gender in juvenile idiopathic arthritis-associated uveitis. J Ophthalmol Published online February 20, 2014. doi:https://doi.org/10.1155/2014/461078
3. Clarke SL, Sen ES, Ramanan AV. Juvenile idiopathic arthritis-associated uveitis. Pediatr Rheumatol Online J 2016;14(1):27.
4. Carvounis PE, Herman DC, Cha S, et al. Incidence and outcomes of uveitis in juvenile rheumatoid arthritis, a synthesis of the literature. Graefes Arch Clin Exp Ophthalmol 2006;244(3):281–290.
5. Heiligenhaus A, Heinz C, Edelsten C, et al. Review for disease of the year: epidemiology of juvenile idiopathic arthritis and its associated uveitis: the probable risk factors. Ocul Immunol Inflamm 2013;21(3):180–191.
6. Sen ES, Ramanan AV. Juvenile idiopathic arthritis-associated uveitis. Clin Immunol 2020;211:108322.
7. Ramanan AV, Guly CM, Keller SY, et al. Clinical effectiveness and safety of baricitinib for the treatment of juvenile idiopathic arthritis-associated uveitis or chronic anterior antinuclear antibody-positive uveitis: study protocol for an open-label, adalimumab active-controlled phase 3 clinical trial (JUVE-BRIGHT). Trials 2021;22(1):689.
8. Constantin T, Foeldvari I, Anton J, et al. Consensus-based recommendations for the management of uveitis associated with juvenile idiopathic arthritis: the SHARE initiative. Ann Rheum Dis 2018;77(8):1107–1117.
9. Foeldvari I, Maccora I, Petrushkin H, et al. New and updated recommendations for the treatment of juvenile idiopathic arthritis–associated uveitis and idiopathic chronic anterior uveitis. Arthritis Care Res (Hoboken) 2023;75(5):975–982.
10. Maccora I, Sen ES, Ramanan AV. Update on noninfectious uveitis in children and its treatment. Curr Opin Rheumatol 2020;32(5):395–402.
11. Ramanan AV, Dick AD, Jones AP, et al; SYCAMORE Study Group. Adalimumab plus methotrexate for uveitis in juvenile idiopathic arthritis. N Engl J Med 2017;376(17):1637–1646.
12. Quartier P, Baptiste A, Despert V, et al; ADJUVITE Study Group. ADJUVITE: a double-blind, randomised, placebo-controlled trial of adalimumab in early onset, chronic, juvenile idiopathic arthritis-associated anterior uveitis. Ann Rheum Dis 2018;77(7):1003–1011.
13. Maccora I, Fusco E, Marrani E, et al. Changing evidence over time: updated meta-analysis regarding anti-TNF efficacy in childhood chronic uveitis. Rheumatology (Oxford) 2021;60(2):568–587.

14. Skrabl-Baumgartner A, Seidel G, Langner-Wegscheider B, et al. Drug monitoring in long-term treatment with adalimumab for juvenile idiopathic arthritis-associated uveitis. Arch Dis Child 2019;104(3):246–250.
15. Breitbach M, Tappeiner C, Böhm MR, et al. Discontinuation of long-term adalimumab treatment in patients with juvenile idiopathic arthritis-associated uveitis. Graefes Arch Clin Exp Ophthalmol 2017; 255(1):171–177.
16. Cecchin V, Zannin ME, Ferrari D, et al. Longterm safety and efficacy of adalimumab and infliximab for uveitis associated with juvenile idiopathic arthritis. J Rheumatol 2018;45(8):1167–1172.
17. Ramanan AV, Dick AD, Guly C, et al; APTITUDE Trial Management Group. Tocilizumab in patients with anti-TNF refractory juvenile idiopathic arthritis-associated uveitis (APTITUDE): a multicentre, single-arm, phase 2 trial. Lancet Rheumatol 2020;2(3):e135–e141.
18. Maccora I, Land P, Miraldi Utz V, et al. Therapeutic potential of JAK inhibitors in juvenile idiopathic arthritis-associated uveitis. Expert Rev Clin Immunol 2023;19(7):689–692.
19. Miserocchi E, Giuffrè C, Cornalba M, et al. JAK inhibitors in refractory juvenile idiopathic arthritis-associated uveitis. Clin Rheumatol 2020; 39(3):847–851.
20. Baquet-Walscheid K, Heinz C, Rath T, et al. Beneficial effect of upadacitinib in an adult patient with juvenile idiopathic arthritis-associated uveitis after unsatisfactory response to tofacitinib: a case report. Ocul Immunol Inflamm 2023;31(5):1079–1080.
21. Bauermann P, Heiligenhaus A, Heinz C. Effect of Janus kinase inhibitor treatment on anterior uveitis and associated macular edema in an adult patient with juvenile idiopathic arthritis. Ocul Immunol Inflamm 2019;27(8):1232–1234.
22. Li Z, Hu J, Bao C, et al. SAT0218 efficacy and safety of baricitinib in mtx-ir patients with rheumatoid arthritis: 52 week results from a phase 3 study (RA-BALANCE). Ann Rheum Dis 2018;77: 969–970.
23. Taylor PC, Keystone EC, van der Heijde D, et al. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. N Engl J Med 2017; 376(7):652–662.
24. Ramanan AV, Quartier P, Okamoto N, et al; JUVE-BASIS investigators; Paediatric Rheumatology International Trials Organisation. Baricitinib in juvenile idiopathic arthritis: an international, phase 3, randomised, double-blind, placebo-controlled, withdrawal, efficacy, and safety trial. Lancet 2023;402(10401):555–570.
25. Giannini, EH, Ruperto, N, Ravelli, A, et al. Preliminary definition of improvement in juvenile arthritis. Arthritis & Rheumatism 1997; 40(7):1202–1209. https://doi.org/10.1002/1529-0131(199707)40:7<1202::aid-art3<3.0.co;2-r.
26. Hampson LV, Whitehead J, Eleftheriou D, et al. Bayesian methods for the design and interpretation of clinical trials in very rare diseases. Stat Med 2014;33(24):4186–4201.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 439–448
DOI 10.1002/acr.25650

American College of Rheumatology
Empowering Rheumatology Professionals

Temporal Trends in and Associations With Nonsteroidal Anti-inflammatory Drug Prescription in Adult and Pediatric Patients With Inflammatory Bowel Disease

Adam S. Mayer,[1,2] Rui Xiao,[1,2] Andrew Grossman,[2] Meenakshi Bewtra,[1] Michael D. George,[1] and Pamela F. Weiss[1,2]

Objective. Recent inflammatory bowel disease (IBD) treatment guidelines have recommended against nonsteroidal anti-inflammatory drug (NSAID) use despite prevalent musculoskeletal symptoms and opioid overuse in this population. Given the discordance between changing national guidelines and potential clinical utility, we sought to assess national temporal trends in prescription NSAID and opioid use for patients with IBD and factors associated with NSAID fill trends.

Methods. This retrospective cohort study of adult and pediatric IBD patients used administrative claims data from 2000 to 2022. Prescription NSAID and opioid fills per calendar year were assessed. Wilcoxon-Cuzick test of trend and generalized estimating equation models evaluated NSAID and opioid fill trends and assessed characteristics associated with NSAID use.

Results. Among the 361,025 IBD patients, there was a significant decreasing trend in the proportion prescribed NSAIDs over time ($P < 0.01$). Fill rates of NSAIDs were markedly lower than opioids across the study period despite an increase in musculoskeletal pain codes. In the multivariable model, opioid prescription (odds ratio [OR] 2.13, 95% confidence interval [CI] 2.11–2.15), a diagnostic code for osteoarthritis (OR 1.57, 95% CI 1.55–1.59), or unspecified joint pain (OR 1.54, 95% CI 1.52–1.56) had strong independent associations with NSAID fill, whereas an age <18 or ≥80 years were associated with significantly lower odds of NSAID fill.

Conclusion. NSAIDs are used by a minority of patients with IBD, with decreasing prescription rates over time despite high rates of opioid use and a doubling of musculoskeletal complaints. NSAID safety needs more thorough examination as an effective and potentially lower-risk analgesic option for patients of all ages with IBD.

INTRODUCTION

Acute and chronic pain are prominent features of inflammatory bowel disease (IBD), which can significantly impact quality of life.[1] The etiology of pain in patients with IBD may be multifactorial but includes inflammatory arthritis or enthesitis in up to 39% of patients.[2] Because of their ease of access, low cost, and effectiveness for a myriad of conditions, including arthritis, nonsteroidal anti-inflammatory drugs (NSAIDs) are an attractive therapeutic option for patients with IBD and musculoskeletal complaints.

However, there is concern that NSAIDs may precipitate intestinal inflammation in patients with IBD. A large 2016 cohort study of patients with self-reported IBD in remission found an increased risk of relapse (risk ratio [RR] 1.65, 95% confidence interval [CI] 1.12–2.44) in those with Crohn disease (CD) subtype who reported at least five NSAID doses per month compared to those with lower intake, although an increased risk associated with acetaminophen exposure was similarly identified.[3] In contrast, a recent systematic review and meta-analysis found no significant increase in risk of IBD exacerbation associated with NSAID exposure (pooled RR 1.29, 95% CI 0.92–1.80). Despite

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (award 5T32-AR-076951-05 to Dr Mayer and award K24-AR-078950 to Dr Weiss) and a Spondyloarthritis Research and Treatment Network Mentored Fellowship Award (to Dr Mayer).

[1]University of Pennsylvania, Philadelphia; [2]Children's Hospital of Philadelphia, Philadelphia, Pennsylvania. Dr Mayer's current address is Center for Discovery and Innovation at Hackensack Meridian Health, Nutley, New Jersey.

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.25650).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25650.

Address correspondence via email to Adam S. Mayer, MD, at adam.mayer@hmh-cdi.org.

Submitted for publication February 19, 2025; accepted in revised form August 29, 2025.

SIGNIFICANCE & INNOVATIONS

- In a large United States cohort of adult and pediatric patients with IBD, we found an overall decreasing trend in prescription NSAID fills over the last two decades despite ongoing high rates of opioid prescription across age groups and a doubling of codes for musculoskeletal pain.
- Several factors independently associated with prescription NSAID use were identified, including codes for osteoarthritis and unspecified joint pain, suggesting that musculoskeletal symptoms are a key driver of NSAID use in patients with IBD.
- Our findings highlight the need to better understand the true risk of NSAIDs for patients with IBD as an effective and potentially lower-risk analgesic option.

mixed data, the most recent 2018 and 2019 American College of Gastroenterology (ACG) treatment guidelines for both CD and ulcerative colitis (UC) strongly recommend against the use of NSAIDs, given the possible risk of intestinal flare.[4,5]

Although the mixed evidence poses a challenge for clinicians caring for adult patients with IBD, even greater challenges exist in the care of pediatric patients with IBD because there are no large studies assessing NSAID risk in this population. Furthermore, the decision to use NSAIDs in patients with IBD is counterweighted by more toxic alternatives including opioids, glucocorticoids, and escalating background immunosuppressive therapy. It is not clear how NSAID prescribing patterns for those with IBD have changed longitudinally in light of sparse and often conflicting evidence and changing national guidelines nor how these trends compare with those of opioids, which are a major alternative analgesic therapy in this population. As such, we sought to evaluate temporal trends in prescription NSAID and opioid fills over time in both adult and pediatric patients with IBD as well as assess factors associated with prescription NSAID use.

MATERIALS AND METHODS

The protocol for the conduct of this study was reviewed by the Children's Hospital of Philadelphia Institutional Review Board and determined to be not human subject research.

Data source. This retrospective cohort study used data from January 2000 to June 2022 in Optum's deidentified Clinformatics Data Mart Database (CDM), a national database of administrative health claims for adult and pediatric members of commercial and Medicare Advantage health plans. Available data included deidentified patient characteristics, International Classification of Diseases (ICD)-9 and ICD-10 diagnosis codes, National Drug Codes (NDCs) for medications, prescription fill dates, pill quantity, and days supplied for a given prescription.

Patient population. Adult (age ≥ 18 years) and pediatric (age < 18 years) patients with IBD were included. A diagnosis of IBD was defined based on validated algorithms[6–11] and required all of the following: (1) at least two inpatient or outpatient diagnosis codes of CD (ICD-9: 556.xx / ICD-10: K51.xx) and/or UC (ICD-9: 555.xx / ICD-10: K50.xx) at least 30 days apart; (2) at least one diagnosis code had to be from an outpatient encounter; and (3) a minimum of six months of continuous enrollment. Select prior claims database studies in pediatric IBD have included less specific codes for indeterminate and noninfective colitis,[12] which we excluded to increase the specificity of our IBD definition and also allow more accurate comparison between the adult and pediatric cohorts by using identical codes for inclusion in both groups. Patients under six years old at the time of meeting IBD criteria were excluded because of the possibility of very-early–onset IBD in this population, a distinct clinical entity from traditional IBD.[13] Patients without pharmacy claims data were also excluded.

The index date was defined as the date the patient met the IBD criteria. All data for outcome analysis started after the index date.

Exposure. The primary exposure was calendar year. Data from each calendar year for a given patient were included if there was at least six months of enrollment data available for that year. Each included calendar year was considered as a separate observation for every patient. Person-months of exposure per calendar year were adjusted for in the primary analysis to account for differential enrollment periods.

Outcomes. The primary outcome was the proportion of patients with more than or equal to one oral NSAID prescription fill each calendar year. Prescription fills for both nonselective and cyclooxygenase-2 (COX-2) selective NSAIDs were assessed (**Supplemental Table** 1). NSAID receipt was dichotomized as ever or never (more than or equal to one NSAID prescription fill) to characterize the overall cohort. NSAID fills by calendar year were used to assess trends in prescription NSAID use over time. A secondary outcome was the proportion of patients with more than or equal to one opioid fill in each calendar year. A sensitivity analysis assessed longitudinal trends of chronic NSAID or opioid use, defined as having at least two NSAID or opioid fills in a calendar year.

Covariates. Potential factors associated with prescription NSAID use included age at the start of calendar year, sex, United States (US) census division (Northeast, South, Midwest, and West), IBD type (CD, UC, or indeterminate colitis [IC]), opioid exposure (or NSAID exposure when assessing opioid fill secondary outcome), nonglucocorticoid immunosuppression regimen (none, conventional synthetic disease-modifying antirheumatic drugs [csDMARD], biologic DMARDs [bDMARDs], or targeted synthetic DMARDs [tsDMARDs]), arthritis, chronic

musculoskeletal pain, depression, anxiety, psychotic disorders, alcohol use, substance use, nicotine/tobacco use, and months enrolled in CDM during eligible calendar years. These covariates were a priori chosen given prior publications demonstrating their association with opioid use and disease severity in adult and pediatric IBD,[12,14–16] which are likely proxies for NSAID exposure. For IBD type, a patient was assigned a diagnosis of CD or UC if ≥75% of IBD codes during the study period were CD- or UC-specific codes, respectively. A patient was labeled as IC if they did not meet this majority threshold for CD or UC codes. Immunosuppression prescription was determined by corresponding NDC or health care common procedure coding system codes. tsDMARDs (ie, Janus kinase inhibitors) were combined with bDMARDs into a category of advanced DMARD therapies because tsDMARDs were first introduced to the US market more than halfway through the study period and used by only a small minority of patients in this study. Arthritis, chronic musculoskeletal pain, mental health disorders, alcohol, substance and nicotine/tobacco use were determined by the presence of more than or equal to corresponding ICD-9 or ICD-10 code (**Supplemental Table** 2) in a given calendar year.

Statistical analysis. Differences in characteristics between (1) adult and pediatric patients with IBD, (2) adult and pediatric patients with IBD and more than or equal to NSAID fill versus none, and (3) those with more than or equal to one NSAID fill during the study period vs those without were summarized using descriptive statistics and compared using Pearson's chi-squared test or *t*-test as appropriate. The Wilcoxon-Cuzick test of trend was used to assess the overall trend in NSAID prescription over time. Generalized estimating equations (GEEs) were used to further determine when adjusting for covariates how the proportion of patients receiving more than or equal to prescription NSAID or more than or equal to opioid during a calendar year changed over time. The GEE model accounted for within-subject correlation as individuals could contribute data for more than one calendar year with time-updating covariates each calendar year. The GEE model was stratified by five age groups (ages 6–17, 18–39, 40–59, 60–79, and 80–99 years). Covariates were included in the multivariable GEE model if there was a significant ($P < 0.05$) association with the primary outcome in univariable models. Linear splines in the GEE models with knots prespecified at calendar years of 2004 and 2018 were used to capture the nonlinear temporal tend. These knots were chosen because of (1) the Federal Drug Administration (FDA) withdrawal of the COX-2 selective inhibitor rofecoxib from the US market in late 2004 and (2) the 2018 ACG (ACG) CD guidelines being the first full-length IBD guidelines to strongly recommend against NSAID use. The slopes of each segment were tested and compared using postestimation commands in Stata. For analysis of opioid fills, a spline regression model was built using a knot at 2012 based on the visual trend. Given that arthritis and chronic musculoskeletal pain are common indications for NSAID prescription in patients with IBD,[17,18] temporal trends of these diagnostic codes were assessed. Finally, because NSAID and opioid fills may be influenced by changing use of advanced DMARD therapies over time, trends in advanced DMARD fills were similarly evaluated. Given the expected sample size of >300,000 patients, there was no anticipation of power limitations for the primary analysis. All statistical analyses were performed using Stata version 18.0.

RESULTS

Cohort characteristics. A flow diagram is shown in Figure 1. A total of 361,025 patients met inclusion criteria,

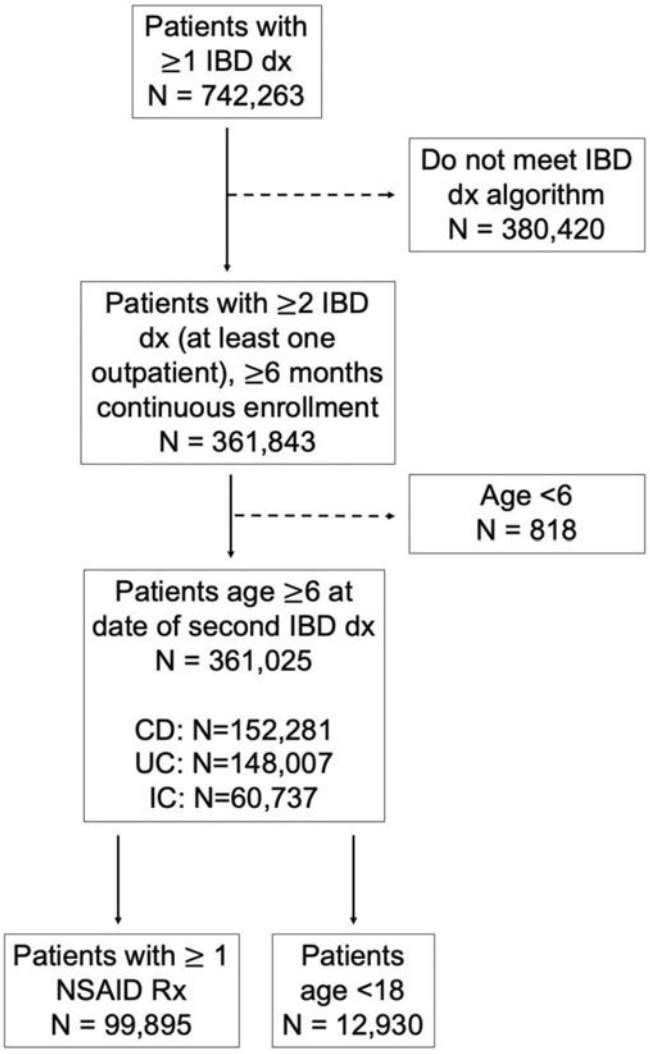

Figure 1. Flow diagram of study cohort. Dashed lines represent excluded patients. CD, Crohn disease; dx, diagnosis; IBD, inflammatory bowel disease; IC, indeterminate colitis; NSAID, nonsteroidal anti-inflammatory drug; Rx, prescription; UC, ulcerative colitis.

including 12,930 (3.6%) patients <18 years old. The cohort characteristics stratified by age group are displayed in **Table** 1. Compared to those <18 years of age, adults with IBD were more likely to be female ($P < 0.01$), have more than or equal to one NSAID fill during the study period (28.5% vs 7.6%; $P < 0.01$) including for nonselective ($P < 0.01$) and COX-2 selective NSAIDs ($P < 0.01$), or have more than or equal to one opioid fill during the study period (54.3% vs 23.5%; $P < 0.01$). Adults were significantly less likely to receive csDMARDs ($P < 0.01$) or bDMARDs/tsDMARDs ($P < 0.01$) than children. Adult patients had a significantly higher prevalence of several comorbidities including inflammatory arthritis (specifically, rheumatoid arthritis and spondylarthritis codes), osteoarthritis, other joint pain not otherwise specified (NOS), and chronic musculoskeletal pain ($P < 0.01$ for all). Similar differences between age groups were followed when assessing the subgroup of those who received more than or equal to one NSAID prescription during the study period, with the exception of spondyloarthritis codes, which did not significantly differ between adult and pediatric patients (13.5% vs 12.8%; $P = 0.34$).

A total of 100,084 (27.7%) patients had at least one prescription NSAID fill during the study period, including 19,359 (5.4%) with at least one COX-2 selective NSAID fill. Of those patients with at least one NSAID fill, the median number of fills was two (interquartile range [IQR] 1–5) over a median of five years of follow-up, the median number of pills supplied per prescription was 66 (IQR 30–210), and the median number of prescription days supplied was 35 (IQR 15–127). The NSAIDs most commonly prescribed included meloxicam (26.4%), celecoxib (18.8%), ibuprofen (15.5%), naproxen (11.8%), and diclofenac (10.1%) (**Supplemental Figure** 1). A total of 192,093 patients (53.2%) had at least one opioid prescription filled. Of

Table 1. Characteristics of patients enrolled, stratified by age group*

	All (n = 361,025)	Age ≥18 years (n = 348,095)	Age <18 years (n = 12,930)
Demographics			
Years enrolled, median (IQR)	3 (2–5)	3 (2–5)	3 (2–5)
Months enrolled per calendar year, median (IQR)	12 (10.5–12)	12 (10.5–12)	12 (10.5–12)
Age at index date, median (IQR)	50 (34–65)	50 (35–65)	14 (12–16)
Female sex, n (%)	197,737 (54.8)	191,982 (55.2)	5,755 (44.5)
Geographic region, n (%)			
Northeast	48,744 (13.5)	47,008 (13.5)	1,736 (13.4)
South	147,837 (40.9)	142,652 (41.0)	5,185 (40.1)
Midwest	96,140 (26.6)	92,268 (26.5)	3,872 (29.9)
West	67,909 (18.8)	65,789 (18.9)	2,120 (16.4)
Unknown	395 (0.1)	378 (0.1)	17 (0.1)
Clinical features, n (%)			
IBD type			
Crohn's disease	152,281 (42.2)	146,892 (42.2)	5,389 (41.7)
Ulcerative colitis	148,007 (41.0)	142,745 (41.0)	5,262 (40.7)
Indeterminate colitis	60,737 (16.8)	58,458 (16.8)	2,279 (17.6)
Inflammatory arthritis	42,848 (11.9)	42,127 (12.1)	721 (5.6)
Rheumatoid arthritis	22,493 (6.2)	22,153 (6.4)	340 (2.6)
Spondyloarthritis	27,352 (7.6)	26,836 (7.7)	516 (4.0)
JIA	392 (0.1)	229 (0.1)	163 (1.3)
Osteoarthritis	113,128 (31.3)	112,942 (31.9)	186 (1.6)
Other joint pain NOS	120,580 (33.4)	119,071 (33.6)	1,509 (12.6)
Chronic musculoskeletal pain	75,082 (20.8)	74,243 (20.9)	839 (7.0)
Depression	108,138 (30.0)	106,328 (30.0)	1,810 (15.2)
Anxiety	111,687 (30.9)	109,596 (30.9)	2,091 (17.5)
Psychotic disorder	20,928 (5.8)	20,676 (5.8)	252 (2.1)
Alcohol use disorder	14,014 (3.9)	13,948 (3.9)	66 (0.6)
Substance use	44,298 (12.3)	44,085 (12.4)	213 (1.8)
Nicotine/tobacco use	95,485 (26.4)	95,365 (26.9)	90 (0.8)
Medications, n (%)			
≥1 NSAID fill	100,084 (27.7)	99,099 (28.5)	985 (7.6)
≥1 COX-2 NSAID fill	19,359 (5.4)	19,189 (5.5)	170 (1.3)
≥1 nonselective NSAID fill	90,316 (25.0)	89,452 (25.7)	864 (6.7)
≥1 opioid fill	192,093 (53.2)	189,056 (54.3)	3,037 (23.5)
Immunosuppression			
None	108,158 (30.0)	106,939 (30.7)	1,219 (9.4)
csDMARD	218,183 (60.4)	209,135 (60.1)	9,048 (70.0)
bDMARD/tsDMARD	107,210 (29.7)	101,876 (29.3)	5,334 (41.3)

* bDMARD, biologic disease-modifying antirheumatic drug; COX-2, cyclooxygenase-2; csDMARD, conventional synthetic DMARD; IBD, inflammatory bowel disease; IQR, interquartile range; JIA, juvenile idiopathic arthritis; NOS, not otherwise specified; NSAID: nonsteroidal anti-inflammatory drug, tsDMARD: targeted DMARD.

those with more than or equal to one opioid fill, the median number of fills was three (IQR 1–10).

Factors associated with prescription NSAID use. In the multivariable model, several variables had a significant and independent association with NSAID prescription (**Table 2**). Patients with an opioid prescription in the same year or diagnoses of inflammatory arthritis, osteoarthritis, or other joint pain NOS were significantly more likely to use NSAIDs. Other clinical characteristics, including diagnoses of chronic musculoskeletal pain, depression, anxiety, or psychotic disorder, had smaller significant effect estimates, whereas IBD type, alcohol use, substance use, and nicotine/tobacco use were not significantly associated with prescription NSAID use. Conversely, male sex, age <18 years or age ≥80 years, and use of advanced immunosuppression (bDMARDs/tsDMARDs alone or in combination with csDMARDs) were independently associated with a lower odds of NSAID prescription. The number of eligible months per calendar year was weakly associated with NSAID prescription.

Temporal trends in prescription NSAID and opioid use. Trends in NSAID prescriptions by calendar year are illustrated in Figure 2A–C. There was a significant decrease in the proportion of patients receiving NSAIDs over time ($P < 0.01$), from 16.0% in 2000 to 6.7% in 2022. There were significant changes in NSAID prescriptions at 2004 and 2018 in the linear spline regression model ($P < 0.01$ for each). Similar spline regressions for the NSAID subtypes revealed a significant decreasing trend in COX-2 selective NSAID fills after 2004 and nonselective NSAID fills after 2018 ($P < 0.01$ for both). These trends were consistent across age groups.

There was a significant downward trend in the proportion of patients with IBD receiving more than or equal to one opioid prescription over time (Figure 2D; $P < 0.01$). The proportion of patients receiving opioid prescriptions was relatively stable before

Table 2. Results of univariable and multivariable GEE regression models for factors associated with NSAID prescription*

Factor	Univariate OR	95% CI	Multivariate OR	95% CI
Demographics				
Age, years, (reference: 18–39)				
6–17	0.49	0.46–0.52	0.54	0.51–0.58
40–59	1.39	1.36–1.41	1.19	1.18–1.21
60–79	1.25	1.23–1.27	1.06	1.04–1.08
80–99	0.73	0.71–0.75	0.67	0.65–0.69
Male sex	0.76	0.75–0.77	0.89	0.88–0.90
Calendar year	0.85	0.85–0.86	0.96	0.96–0.96
Number of eligible months per calendar year	1.09	1.09–1.09	1.04	1.04–1.04
Years enrolled in CDM	1.00	1.00–1.00	0.99	0.99–1.00
Geographic region (reference: northeast)				
South	1.29	1.27–1.32	1.12	1.10–1.14
Midwest	1.00	0.97–1.02	0.91	0.89–0.93
West	0.93	0.90–0.95	0.88	0.71–1.13
Clinical features				
IBD type (reference: CD)				
UC	1.01	1.00–1.02	1.00	0.99–1.02
IC	1.03	1.01–1.05	0.99	0.98–1.01
≥1 Opioid fill in same year	2.64	2.62–2.67	2.13	2.11–2.15
Immunosuppression (reference: none)				
csDMARD	1.01	1.00–1.03	1.00	0.99–1.01
bDMARD/tsDMARD	0.84	0.82–0.86	0.88	0.87–0.89
csDMARD+bDMARD/tsDMARD	0.92	0.91–0.94	0.90	0.88–0.91
Inflammatory arthritis	1.83	1.80–1.86	1.25	1.23–1.28
Osteoarthritis	2.10	2.08–2.12	1.57	1.55–1.59
Other joint pain NOS	1.63	1.61–1.65	1.54	1.52–1.56
Chronic musculoskeletal pain	1.67	1.65–1.69	1.11	1.10–1.12
Depression	1.32	1.31–1.34	1.03	1.02–1.04
Anxiety	1.27	1.26–1.29	1.04	1.02–1.06
Psychotic disorders	1.44	1.41–1.48	1.10	1.07–1.13
Substance use	1.33	1.31–1.36	1.01	0.98–1.02
Alcohol use	1.22	1.18–1.27	1.01	0.98–1.05
Nicotine/tobacco use	1.31	1.29–1.32	1.00	0.99–1.02

* bDMARD, biologic disease-modifying antirheumatic drug; CD, Crohn disease; CDM, Clinformatics Data Mart Database; CI, confidence interval; csDMARD, conventional synthetic DMARD; GEE, generalized estimating equation; IBD, inflammatory bowel disease; IC, indeterminate colitis; NOS, not otherwise specified; NSAID, nonsteroidal anti-inflammatory drug; OR, odds ratio; tsDMARD, targeted synthetic DMARD; UC, ulcerative colitis.

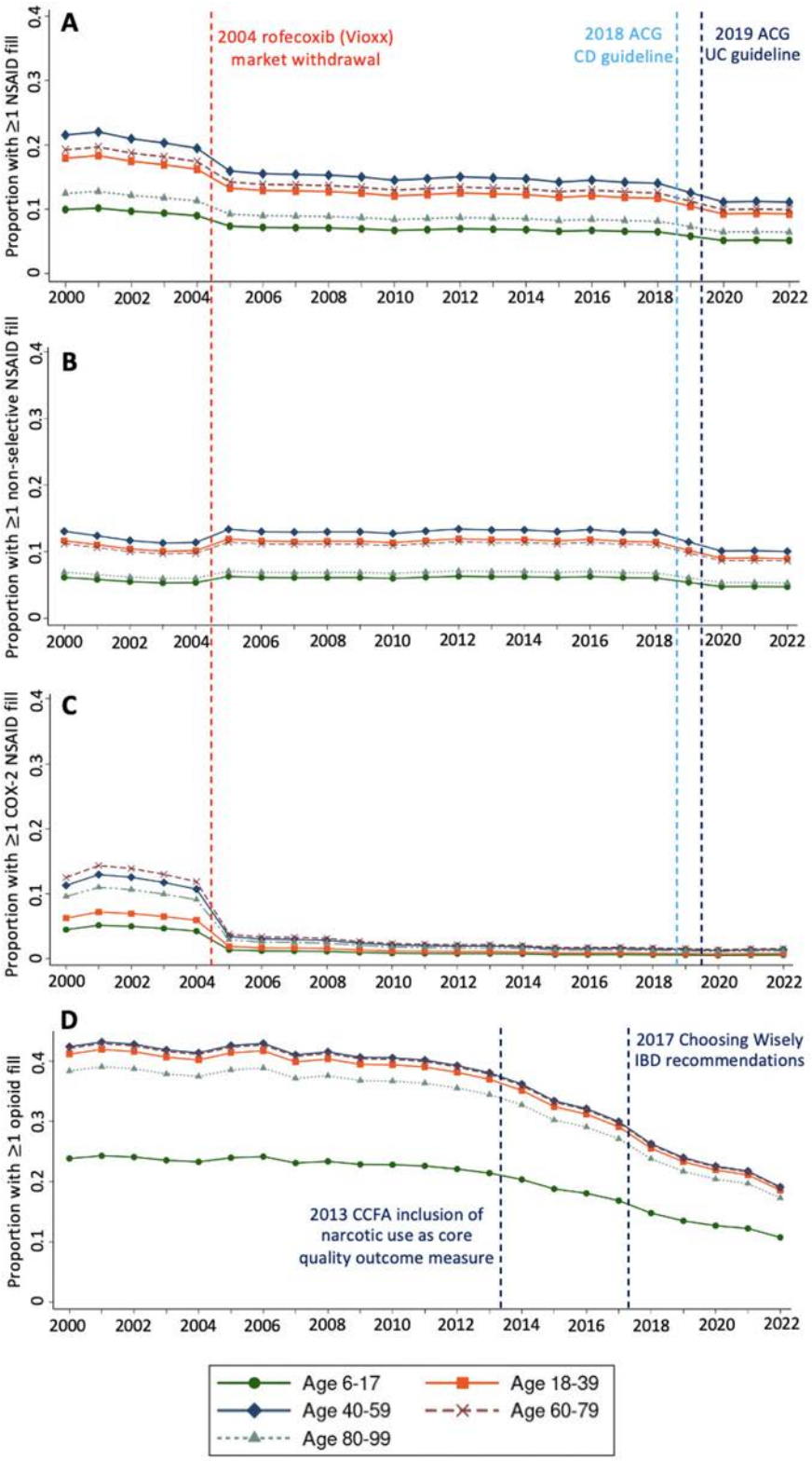

Figure 2. Longitudinal trends in (A) prescription NSAID, (B) nonselective NSAID, (C) COX-2 selective NSAID, and (D) opioid fills. Annual proportions of prescriptions fills are estimated from the generalized estimating equation model adjusted for covariates. Relevant gastroenterological guidelines advising against NSAID use in IBD are indicated by the vertical dashed lines in (A)–(C). Relevant national public health and quality of care guidelines for opioid use in IBD are indicated by the vertical dashed lines in (D). ACG, American College of Gastroenterology; AGA, American Gastroenterological Association; CCFA, Crohn's & Colitis Foundation of America; CD, Crohn disease; COX-2, cyclooxygenase-2; IBD, inflammatory bowel disease; NSAID, nonsteroidal anti-inflammatory drug; UC, ulcerative colitis.

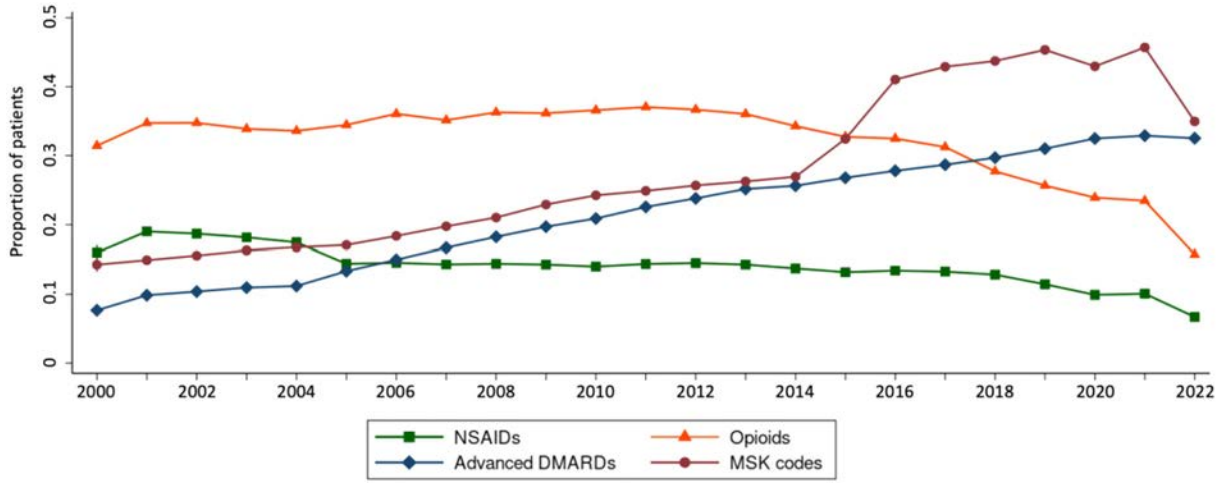

Figure 3. Comparative trends in musculoskeletal symptom codes and prescription NSAID, opioid, and advanced DMARD (includes biologic and targeted synthetic DMARDs) fills. MSK codes refer to diagnosis codes for arthritis or chronic musculoskeletal pain. DMARD: disease modifying antirheumatic drug; MSK: musculoskeletal; NSAID, nonsteroidal anti-inflammatory drug. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25650/abstract.

2012 ($P = 0.89$), but there was a significant downslope after 2012 ($P < 0.01$), with 15.9% of enrolled patients prescribed opioids in 2022. Despite this downward trend, opioid use remained nearly twice as common as NSAID use across all age groups throughout the study period. In contrast to the decreasing trend in NSAID and opioid use over time, the proportion of patients with advanced DMARD fills and codes for arthritis or chronic musculoskeletal pain in a given calendar year significantly increased over time ($P < 0.01$), with musculoskeletal symptom codes increasing from 14.3% in 2000 to 35.0% in 2022 (Figure 3). The longitudinal pattern of NSAID and opioid fills in the primary analysis did not differ across IBD subtypes or among chronic users (**Supplemental Figures** 2 and 3).

DISCUSSION

Leveraging a large cohort of adult and pediatric patients with IBD from a national health care administrative claims database, we found an overall decreasing trend in the proportion of patients receiving prescription NSAIDs over time in all age groups, including (1) a marked sustained decrease in COX-2 selective inhibitors after 2004, when rofecoxib was withdrawn from the US market, and (2) a further decrease in all prescription NSAID fills after the most recent ACG IBD treatment guideline recommendations. The proportion of patients receiving prescription NSAIDs remained well below that of those receiving opioids across all calendar years irrespective of age group, despite a more-than--doubling of codes for musculoskeletal pain over time. Several factors strongly associated with NSAID use were identified including opioid prescription or codes for inflammatory arthritis, osteoarthritis, or joint pain NOS in the same calendar year. Pediatric patients had the lowest odds of NSAID prescription. There are several strengths to our study. The inclusion of a pediatric cohort is novel because to the best of our knowledge there are no published studies assessing trends in NSAID prescription in pediatric IBD and particularly how this directly compares to adult cohorts. Our study spanned 22 years and thus allowed for the determination of longer-term comparative trends in NSAID and opioid fills across changing guidelines and a dynamic landscape of novel therapies for the treatment of IBD and ultimately highlight the need for further study of NSAID safety as an analgesic that may be less toxic than opioids in this population.

The decreasing trend in prescription NSAID use over time in our US-based cohort is consistent with that reported previously in Danish patients with IBD[19] as well as longitudinal NSAID prescription trends in other diseases such as axial spondyloarthritis and cardiovascular disease (CVD).[20–22] The precipitous drop in COX-2 prescription fills after the FDA withdrawal of rofecoxib in 2004 is a well-established phenomenon and there appears to have been a contemporaneous compensatory increase in nonselective NSAID prescriptions, which has been previously demonstrated outside of IBD.[23] Subsequent to 2004, the proportion of patients using COX-2 selective NSAIDs remained markedly lower than nonselective NSAIDs, despite data suggesting relative safety in patients with IBD.[24] However, there was also a significant decrease in the proportion of patients with nonselective NSAID prescriptions after 2018. We hypothesize this may be due to the 2018 and 2019 ACG guidelines[4,5] strongly recommending against the use of NSAIDs in patients with IBD, which were the first full-length IBD guidelines to do so and thus highlighting the impact of national treatment guidelines on provider prescribing practices. As supported by the multivariable model, the advent of more effective immunosuppressive medications over time (eg, bDMARDs and tsDMARDs) is another factor likely contributing to the decreasing rate of prescription NSAID use due to better control of underlying IBD and associated musculoskeletal

disease, but there is still a decreasing trend in NSAID prescription independent of immunosuppressive regimen. More recent treatment protocols emphasizing earlier treatment of IBD with biologic and small molecule therapies may also contribute to this trend. The pattern of NSAID use over time in children with IBD mirrored that of adult patients but with a lower proportion of patients with fills each year for each NSAID subtype and with a notably low prevalence of COX-2 selective NSAID use overall. The lower proportion of NSAID prescriptions may be related to the earlier and more frequent use of biologic therapies observed in the pediatric group, which may result in lower disease activity, as well as the lower prevalence of comorbidities relative to adults, including arthritis and chronic musculoskeletal pain.[25]

The proportion of patients receiving opioids decreased over time after 2012, consistent with several prior studies in adult and pediatric IBD.[12,26–28] This trend may be due in large part to national shifts in medical practice and increased awareness of risks associated with opioid use.[29–31] These shifts in practice are also reflected in the inclusion of narcotic medication use as a core outcome measure in the quality indicators for IBD from the Crohn's & Colitis Foundation of America in 2013[32] as well as the 2017 recommendation to avoid long-term opioid therapy for abdominal pain in IBD from the Choosing Wisely campaign.[33] Nevertheless, opioid fills were nearly twice as common as NSAID fills in our cohort, and the prevalence of opioid fills remained markedly higher than that of NSAIDs across the entire 22 year study period, acknowledging that data on over-the-counter NSAID use are missing in claims databases, whereas opioids are only obtained by prescription. These findings are consistent with a large 2014 Canadian retrospective cohort study showing that 53% of patients with IBD will have opioid exposure within five years of diagnosis[34] but further illustrate how these high rates of opioid prescription have persisted over time. The continued high rates of opioid use are in the context of increasing rates of musculoskeletal symptoms seen in our study, which are consistent with published global trends and may be related to increased provider awareness and an increase in the median age of the US population over time.[35,36] The increased specificity of musculoskeletal symptom codes in the transition from ICD-9 to ICD-10 in 2015 likely explains the significant increase in musculoskeletal symptom codes observed in that year,[37] but there is still an overall increase in the rates of musculoskeletal codes across the study independent of the ICD transition period. Ultimately, the pervasive use of opioids in both adults and children, along with increasing rates of musculoskeletal pain, highlight the ongoing challenge to identify ways of providing effective, yet safe, analgesia in IBD across the age spectrum.

Several factors were found to be significantly associated with prescription NSAID use. Opioid prescription had the strongest independent association because patients receiving opioids are more likely to have pain and severe disease[15,16] and thus more likely to receive other analgesics. Inflammatory arthritis, osteoarthritis, other joint pain NOS, and chronic musculoskeletal pain were all associated with an increased odds of NSAID fill because NSAIDs are commonly prescribed for these indications. In particular, NSAIDs are considered first-line therapy for several musculoskeletal conditions associated with IBD, such as enthesitis, axial spondyloarthritis, and psoriatic arthritis,[38,39] highlighting the relevance of this drug class in specific subpopulations of patients with IBD as a potentially safer alternative analgesic. Male sex was independently associated with a lower odds of NSAID prescription fill, possibly because of the greater burden of comorbidities in male patients that are relative contraindications to NSAID use (eg, CVD). Furthermore, women with IBD and spondyloarthritis have been shown to have poorer quality of life metrics and increased self-reported pain compared to men, which may also influence NSAID prescription.[40,41] Age <18 years, as discussed previously, and age ≥80 years were associated with a lower odds of NSAID use. Elderly patients are more likely to have comorbidities that may preclude NSAID use and are known to have higher risks associated with exposure to this drug class.[42] Interestingly, although several covariates in our study such as depression, anxiety, and substance use have been previously reported to be associated with opioid prescription,[12,14–16] they did not have particularly strong associations with NSAID use in our cohort.

There are several limitations to this study that should be acknowledged. Prescription fills may not always translate to actual patient use, which would bias toward overestimating NSAID use, although this overestimation is likely outweighed by an inability to assess over-the-counter NSAID use, leading to an underestimation of total NSAID use. However, this study is specifically evaluating patterns of prescription NSAID fills and how these patterns have changed over time. Furthermore, given recent guidelines making strong recommendations against NSAID use in IBD, patients likely are frequently counseled on avoiding over-the-counter NSAIDs. There is a potential of misclassification of patients as having IBD when using claims data, but validated algorithms were adapted using multiple IBD codes separated over time. Patients had to be enrolled for at least six months in a calendar year for data inclusion, but it is possible those enrolled for longer periods in a given year (ie, 12 months) would have a greater potential for NSAID exposure than those enrolled for a shorter period. The number of eligible months per calendar year was therefore adjusted for in our multivariable model. Although data were lacking on important factors that may serve as unmeasured confounders, such as IBD disease severity, codes for covariates that are important indications for NSAID prescription in IBD were included, namely arthritis and chronic musculoskeletal pain but without the ability to confirm such diagnoses when using claims data. Data on prescriber specialty and surgical codes were not assessed, and thus the association of these factors with prescription NSAID fills was not ascertained. There is a possibility of missing data if certain diagnoses were not coded

for during insurance billing, but this issue is inherent to retrospective administrative data and missingness is expected to be nondifferential between NSAID exposure groups. Finally, although our study did not include uninsured individuals, our findings in this large national cohort are likely generalizable to other commercially insured patients with IBD in the US.

In conclusion, prescription NSAIDs are used in a minority of patients of all ages with IBD, but prescription fill rates have decreased over time, particularly for COX-2 selective NSAIDs after 2004 and more generally after the recent ACG IBD treatment guidelines in 2018 and 2019. Pediatric patients were much less likely to receive prescription NSAIDs than adults, even though the side effect profile may be more favorable in this population. Although greater use of biologic/tsDMARDs may be partially responsible for lower rates of NSAIDs, the continued high rates of opioid use and coding for musculoskeletal complaints demonstrate an important remaining gap in pain treatment in this population. There is a need to better evaluate NSAID safety in both adults and children with IBD to determine whether NSAIDs can be an effective and potentially lower-risk analgesic option than opioids for these patients.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Mayer 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. Lix LM, Graff LA, Walker JR, et al. Longitudinal study of quality of life and psychological functioning for active, fluctuating, and inactive disease patterns in inflammatory bowel disease. Inflamm Bowel Dis 2008;14(11):1575–1584.
2. de Vlam K, Mielants H, Cuvelier C, et al. Spondyloarthropathy is underestimated in inflammatory bowel disease: prevalence and HLA association. J Rheumatol 2000;27(12):2860–2865.
3. Long MD, Kappelman MD, Martin CF, et al. Role of nonsteroidal anti-inflammatory drugs in exacerbations of inflammatory bowel disease. J Clin Gastroenterol 2016;50(2):152–156.
4. Rubin DT, Ananthakrishnan AN, Siegel CA, et al. ACG clinical guideline: ulcerative colitis in adults. Am J Gastroenterol 2019;114(3):384–413.
5. Lichtenstein GR, Loftus EV, Isaacs KL, et al. ACG clinical guideline: management of Crohn's disease in adults. Am J Gastroenterol 2018;113(4):481–517.
6. Liu L, Allison JE, Herrinton LJ. Validity of computerized diagnoses, procedures, and drugs for inflammatory bowel disease in a northern California managed care organization. Pharmacoepidemiol Drug Saf 2009;18(11):1086–1093.
7. Hou JK, Tan M, Stidham RW, et al. Accuracy of diagnostic codes for identifying patients with ulcerative colitis and Crohn's disease in the Veterans Affairs Health Care System. Dig Dis Sci 2014;59(10):2406–2410.
8. Ye Y, Manne S, Bennett D. Identifying patients with inflammatory bowel diseases in an administrative health claims database: do algorithms generate similar findings? Inquiry. 2019;56:46958019887816.
9. McAuliffe ME, Lanes S, Leach T, et al. Occurrence of adverse events among patients with inflammatory bowel disease in the HealthCore Integrated Research Database. Curr Med Res Opin 2015;31(9):1655–1664.
10. Waljee AK, Wiitala WL, Govani S, et al. Corticosteroid use and complications in a US inflammatory bowel disease cohort. PLoS One 2016;11(6):e0158017.
11. Gan RW, Sun D, Tatro AR, et al. Replicating prediction algorithms for hospitalization and corticosteroid use in patients with inflammatory bowel disease. PLoS One 2021;16(9):e0257520.
12. Bilgic Dagci AO, Chang JC, Xiao R, et al. Opioid use in children with inflammatory bowel disease-related arthritis. Clin Exp Rheumatol 2023;41(7):1553–1560.
13. Uhlig HH, Schwerd T, Koletzko S, et al; COLORS in IBD Study Group and NEOPICS. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology 2014;147(5):990–1007.e3.
14. Niccum B, Moninuola O, Miller K, et al. Opioid use among patients with inflammatory bowel disease: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2021;19(5):895–907.e4.
15. Cross RK, Wilson KT, Binion DG. Narcotic use in patients with Crohn's disease. Am J Gastroenterol 2005;100(10):2225–2229.
16. Hanson KA, Loftus EV Jr, Harmsen WS, et al. Clinical features and outcome of patients with inflammatory bowel disease who use narcotics: a case-control study. Inflamm Bowel Dis 2009;15(5):772–777.
17. Felder JB, Korelitz BI, Rajapakse R, et al. Effects of nonsteroidal anti-inflammatory drugs on inflammatory bowel disease: a case-control study. Am J Gastroenterol 2000;95(8):1949–1954.
18. Gleeson MH, Davis AJM. Non-steroidal anti-inflammatory drugs, aspirin and newly diagnosed colitis: a case-control study. Aliment Pharmacol Ther 2003;17(6):817–825.
19. Dalal RS, Nørgård BM, Zegers FD, et al. Older adult-onset of inflammatory bowel diseases is associated with higher utilization of analgesics: a nationwide cohort study. Am J Gastroenterol 2024;119(2):323–330.
20. Hwang AY, Smith SMUS. U.S. trends in prescription nonsteroidal anti-inflammatory drug use among patients with cardiovascular disease, 1988-2016. Pharmacotherapy 2021;41(3):247–256.
21. Degirmenci B, Peloquin C, Lodi S, et al. Opioid prescription rates in patients with radiographic axial spondyloarthritis in the US between 2008-2021: a Joinpoint regression analysis [abstract]. Arthritis Rheumatol 2024;76(suppl 9). https://acrabstracts.org/abstract/opioid-prescription-rates-in-patients-with-radiographic-axial-spondyloarthritis-in-the-us-between-2008-2021-a-joinpoint-regression-analysis/.
22. Sheahan A, Anjohrin S, Suruki R, et al. Opioid use surrounding diagnosis and follow-up in patients with ankylosing spondylitis, psoriatic arthritis, and rheumatoid arthritis: results from US claims databases. Clin Rheumatol 2024;43(6):1897–1907.
23. Sun SX, Lee KY, Bertram CT, et al. Withdrawal of COX-2 selective inhibitors rofecoxib and valdecoxib: impact on NSAID and gastroprotective drug prescribing and utilization. Curr Med Res Opin 2007;23(8):1859–1866.
24. El Miedany Y, Youssef S, Ahmed I, et al The gastrointestinal safety and effect on disease activity of etoricoxib, a selective COX-2 inhibitor in inflammatory bowel diseases. Am J Gastroenterol 2006;101(2):311–317.

25. Rosenbloom BN, Rabbitts JA, Palermo TM. A developmental perspective on the impact of chronic pain in late adolescence and early adulthood: implications for assessment and intervention. Pain 2017; 158(9):1629–1632.
26. Chhibba T, Guizzetti L, Seow CH, et al. Frequency of opioid prescription at emergency department discharge in patients with inflammatory bowel disease: a nationwide analysis. Clin Gastroenterol Hepatol 2021;19(10):2064–2071.e1.
27. Love B, Finney C, Gaidos J. P133 prescription opioid use among veterans with inflammatory bowel disease. Gastroenterology 2020; 158(3):S43–S44.
28. Wren AA, Bensen R, Sceats L, et al. Starting young: trends in opioid therapy among us adolescents and young adults with inflammatory bowel disease in the Truven MarketScan Database between 2007 and 2015. Inflamm Bowel Dis 2018;24(10):2093–2103.
29. Schieber LZ, Guy GP Jr, Seth P, et al. Trends and patterns of geographic variation in opioid prescribing practices by state, United States, 2006-2017. JAMA Netw Open 2019;2(3):e190665.
30. Sun EC, Darnall BD, Baker LC, et al. Incidence of and risk factors for chronic opioid use among opioid-naive patients in the postoperative period. JAMA Intern Med 2016;176(9):1286–1293.
31. Renny MH, Yin HS, Jent V, et al. Temporal trends in opioid prescribing practices in children, adolescents, and younger adults in the US from 2006 to 2018. JAMA Pediatr 2021;175(10):1043–1052.
32. Melmed GY, Siegel CA, Spiegel BM, et al. Quality indicators for inflammatory bowel disease: development of process and outcome measures. Inflamm Bowel Dis 2013;19(3):662–668.
33. Nguyen GC, Boland K, Afif W, et al. Modified Delphi process for the development of choosing wisely for inflammatory bowel disease. Inflamm Bowel Dis 2017;23(6):858–865.
34. Targownik LE, Nugent Z, Singh H, et al. The prevalence and predictors of opioid use in inflammatory bowel disease: a population-based analysis. Am J Gastroenterol 2014;109(10):1613–1620.
35. Wilder K, Mackun P. Median age in 192 metro areas higher than national median of 39.1. Accessed August 12, 2025. https://www.census.gov/library/stories/2025/06/metro-areas-median-age.html
36. 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.
37. Coleman BC, Goulet JL, Higgins DM, et al. ICD-10 coding of musculoskeletal conditions in the Veterans Health Administration. Pain Med 2021;22(11):2597–2603.
38. Ward MM, Deodhar A, Gensler LS, et al. 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network Recommendations for the Treatment of Ankylosing Spondylitis and Nonradiographic Axial Spondyloarthritis. Arthritis Care Res (Hoboken) 2019;71(10):1285–1299.
39. Coates LC, Soriano ER, Corp N, et al; GRAPPA Treatment Recommendations domain subcommittees. Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA): updated treatment recommendations for psoriatic arthritis 2021. Nat Rev Rheumatol 2022;18(8):465–479.
40. Tournadre A, Pereira B, Lhoste A, et al. Differences between women and men with recent-onset axial spondyloarthritis: results from a prospective multicenter French cohort. Arthritis Care Res (Hoboken) 2013;65(9):1482–1489.
41. Casellas F, López-Vivancos J, Vergara M, et al. Impact of inflammatory bowel disease on health-related quality of life. Dig Dis 1999; 17(4):208–218.
42. Wehling M. Non-steroidal anti-inflammatory drug use in chronic pain conditions with special emphasis on the elderly and patients with relevant comorbidities: management and mitigation of risks and adverse effects. Eur J Clin Pharmacol 2014;70(10):1159–1172.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 449–455
DOI 10.1002/acr.25643

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Gastrointestinal Impact and Tool Performance in Juvenile Systemic Sclerosis Using the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract, Version 2.0 Assessment

Sophie Stefancic,[1] Amanda Robinson,[2,3] Haley J. Havrilla,[2] Vibha Sood,[1,2] and Kathryn S. Torok[1,2]

Objective. The objective of this study is to characterize gastrointestinal (GI) manifestations in patients with juvenile-onset systemic sclerosis (jSSc) using the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract, version 2.0 (UCLA GIT 2.0) patient-reported outcome (PRO) instrument, and to evaluate its validity and responsiveness in this population.

Methods. Patients with jSSc from the National Registry for Childhood Onset Scleroderma who completed the UCLA GIT 2.0 were included. Demographic and clinical data, domain, and total UCLA GIT 2.0 scores were summarized. Convergent validity was assessed by Spearman correlations with the Scleroderma Health Assessment Questionnaire GI visual analog scale (SHAQ-GI-VAS) and the SHAQ Disease VAS (SHAQ-DIS-VAS). Responsiveness was explored in patients with paired UCLA GIT 2.0 assessments one year later.

Results. Fifty-one patients with jSSc (mean age at onset, 9.8 years; mean disease duration, 4.4 years) had a mean UCLA GIT 2.0 total score of 0.30, indicating a mild GI burden. Distension and bloating and reflux were the most affected domains, each reported in more than 70% of patients. Total and subscale UCLA GIT 2.0 scores showed moderate to strong significant correlations with the SHAQ-GI-VAS and SHAQ-DIS-VAS, supporting convergent validity. Among 22 patients with paired data, the mean total UCLA GIT 2.0 improved by 0.11 points (P = 0.039), and 55% of these patients achieved a clinically important improvement in at least one domain, indicating preliminary responsiveness.

Conclusion. The UCLA GIT 2.0 captures the frequency and severity of GI symptoms in patients with jSSc and demonstrates acceptable validity and sensitivity to change. Although developed for adults, the instrument appears suitable for monitoring GI outcomes in pediatric patients with SSc in both research and clinical settings.

INTRODUCTION

Systemic sclerosis (SSc) is a complex autoimmune disease characterized by immune dysregulation, fibrosis, and vascular abnormalities, resulting in multisystem involvement. Among the affected organs, the gastrointestinal (GI) tract is one of the most significantly impacted in both patients with adult-onset SSc and juvenile-onset systemic sclerosis (jSSc).[1,2] GI manifestations in patients with SSc occur throughout the GI tract and can include dry mouth, esophageal dysmotility, delayed gastric emptying, small bowel bacterial intestinal overgrowth, malabsorption, constipation, and fecal incontinence.[2] Although these symptoms are well-documented in adults, fewer studies describe GI manifestations in patients with jSSc.[1,3–5]

A recent Childhood Arthritis and Research Alliance (CARRA) cohort study of 64 patients with jSSc identified that GI involvement negatively impacted patient-reported quality of life (QoL) and patient-reported global wellbeing and was linked to physician-reported classification of global functional status using the American College of Rheumatology (ACR) criteria.[1] Despite this, GI involvement in patients with jSSc remains underrecognized and poorly understood. This may be because of

Supported by The Nancy Taylor Foundation for Chronic Diseases Inc (EIN 26-2343375).

[1]University of Pittsburgh, Pittsburgh, Pennsylvania; [2]University of Pittsburgh Medical Center, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania; [3]University of Utah, Salt Lake City.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25643.

Address correspondence via email to Kathryn S. Torok, MD, at kathryn.torok@chp.edu.

Submitted for publication February 16, 2025; accepted in revised form August 18, 2025.

SIGNIFICANCE & INNOVATIONS

- The University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract, version 2.0 (UCLA GIT 2.0) instrument performed well for patients with juvenile systemic sclerosis (jSSc), with total and domain subscale scores comparable with patients with adult SSc, demonstrating face and content validity and reinforcing its suitability in pediatric populations. Strong correlations with the Scleroderma Health Assessment Questionnaire GI visual analog scale and Scleroderma Health Assessment Questionnaire Disease visual analog scale support its convergent construct validity.
- Distension, bloating, and reflux were the most affected domains in patients with jSSc, consistent with data from adults with SSc. These findings highlight the need for targeted supportive and pharmacological interventions for children with jSSc.
- Pediatric patients often underreport GI symptoms to physicians, emphasizing the importance of incorporating a validated patient-reported outcome instrument like the UCLA GIT 2.0 into routine clinical care to enhance symptom detection and management in patients with jSSc.

communication challenges between younger patients and physicians, as well as the subtlety of symptoms like decreased appetite, weight loss, decreased body mass index (BMI), and failure to thrive,[6] which may be overlooked compared to more overt complaints of abdominal pain. Children and adolescents often struggle to articulate GI issues, suggesting that patient-reported outcomes (PROs) may be more effective in capturing GI symptoms than traditional physician questioning.[7]

In adult patients with SSc, GI involvement is significant, affecting up to 90% of patients,[2] leading to the development of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract, version 2.0 (UCLA GIT 2.0) instrument.[8] This validated PRO instrument assesses GI involvement and its impact on QoL and was developed with patient input. Several studies have demonstrated its usefulness for adult patients with SSc for accurately capturing GI symptoms,[9,10] correlating well with other measures such as the Scleroderma Health Assessment Questionnaire visual analog scale (SHAQ-VAS).[11] Given its success in adults, it is logical to explore the use of the UCLA GIT 2.0 in capturing the GI symptom burden in patients with jSSc. In this study, we evaluate both the GI symptom burden in patients with jSSc and the performance of the UCLA GIT 2.0 in this pediatric population, including its domain structure, construct validity, and preliminary responsiveness to clinical change.

PATIENTS AND METHODS

Patient population and measures collected. Participants were consecutively enrolled in the National Registry for Childhood Onset Scleroderma (IRB PRO11060222), a prospective observational registry of patients with jSSc evaluated at a single tertiary care multidisciplinary center. Patients who fulfilled the 2013 ACR/EULAR criteria for SSc[12] and completed the UCLA GIT 2.0 at a clinical study visit between July 2018 to January 2023 were included. Of 98 patients enrolled in the registry by January 10, 2023, 57 had a UCLA GIT 2.0 recorded, and 51 had fully completed and scorable surveys for inclusion in the primary cross-sectional analysis. Demographic and clinical characteristics were collected, including age, sex, disease subtype, autoantibody status, medications, SSc-related clinical variables (eg, modified Rodnan Skin Score [mRSS]), GI and pulmonary evaluations, and PROs were collected. Summary statistics were then performed for all variables collected at the study visit, defined as the first clinical encounter at which the UCLA GIT 2.0 was completed, in tandem with routinely collected clinical research forms. To evaluate construct validity, cross-sectional correlations were assessed between UCLA GIT 2.0 scores and two concurrent SHAQ-VAS measures. For a subset of patients (n = 22) with a second UCLA GIT 2.0 completed approximately one year later (± 6 months), the initial visit was treated as a baseline for paired longitudinal comparison to explore preliminary responsiveness.

UCLA GIT 2.0. The UCLA GIT 2.0 comprises 34 questions across six subscales: reflux, distension and bloating, fecal soilage, diarrhea, social functioning, and emotional wellbeing.[8] Patients select a response for each question based on the frequency in the past week (no days, 1–2 days, 3–4 days, and 5–7 days), scored from 0 to 3, with 3 indicating the highest burden affecting QoL. Diarrhea and constipation subscales use different scoring ranges: diarrhea (0–2) and constipation (0–2.5), incorporating yes/no (measured as 1/0, respectively) questions. Some classify the severity of the UCLA GIT 2.0 into mild, moderate, and severe with scores of 0.0 to 0.49, 0.50 to 1.0, and 1.1 to 3.0, respectively.[8] The UCLA GIT 2.0 was administered to children aged five to seven with guardian assistance, whereas those aged eight and older completed it independently.

SHAQ-VAS. PROs collected concurrently with the UCLA GIT 2.0 included the Scleroderma Health Assessment Questionnaire GI visual analog scale (SHAQ-GI-VAS), and the SHAQ disease visual analog scale (SHAQ-DIS-VAS). The SHAQ-GI-VAS, a validated general GI PRO used with patients with SSc, was included to assess convergent construct validity with the UCLA GIT 2.0, consistent with prior studies among adult patients with SSc.[8,11] The SHAQ-GI-VAS asks, "In the past week, how much did your GI symptoms interfere with function?" The SHAQ-DIS-VAS assesses overall disease severity with the question, "Overall, considering how much pain, discomfort, limitations in your daily life and other changes in your body and life, how severe would you rate your scleroderma today?" The SHAQ-DIS-VAS was used in analysis to reflect prior jSSc studies showing that GI involvement

had the highest impact on overall QoL and wellbeing.[1] Responses for both are marked on a 100mm line, with higher scores indicating more severe disease or symptom burden (0–3 scale).[13]

Statistical analysis. Summary statistics were used to analyze clinical variables, UCLA GIT 2.0 total and subdomain scores, as well as the SHAQ-GI-VAS and SHAQ-DIS-VAS scores at the concurrent clinical visit. Spearman's correlation coefficients were used to assess convergent construct validity between the UCLA GIT 2.0 and SHAQ-GI-VAS and SHAQ-DIS-VAS. Correlation strengths were categorized as weak ($r_s < 0.4$), moderate (r_s 0.4–0.7), and strong ($r_s > 0.7$). For exploratory longitudinal analysis, paired UCLA GIT 2.0 data from patients who completed the survey and again at 1 year ± 6 months after first collection were analyzed. Before applying a paired *t*-test, the distribution of within-participant differences was assessed using Q–Q plots and the Shapiro–Wilk test to confirm approximate normality. A paired *t*-test was then used to assess mean changes in UCLA GIT 2.0 Total score over time. Cohen's *d* for paired samples was calculated to estimate effect size. A two-sided *P* value < 0.05 was considered statistically significant. Descriptive and correlation analyses were performed using Microsoft Excel and Minitab Statistical Software. Q–Q plots, normality testing, and effect size calculations were performed using Python 3.10 with SciPy and Matplotlib libraries.

RESULTS

Patient demographic and clinical characteristics. A total of 51 patients with jSSc completed the UCLA GIT 2.0 assessment at their clinical cohort visit and were included in the primary cross-sectional study. For 33 patients (65%), this was their first jSSc visit to our scleroderma center and date of entry into the cohort registry. Demographics, subtypes, autoantibody profiles, and clinical variables are summarized in Table 1. The mean age of disease onset was 9.8 ± 4.3 years, with a mean disease duration of 4.4 ± 3.3 years at the time of the initial UCLA GIT 2.0 assessment, resulting in a mean age of 14 ± 4.6 years at the study visit. The majority of patients were female (71%) and White (69%). The most common subtype was diffuse cutaneous SSc (47%). Autoantibody testing revealed that Scl-70 (31%) and PM-Scl (20%) levels were the most frequently identified. The mean mRSS was 8 ± 10.6 (range, 0–44).

GI involvement was present in 77% of patients (n = 39/51), with 73% (n = 37) showing abnormalities on GI studies performed up to the time of study visit (Table 1). Abnormal GI studies were defined as the presence of clinically significant findings on modified barium swallow study, gastric emptying studies, manometry, pH probe testing, or endoscopy, based on the interpreting pediatric gastroenterologist's or radiologist's report at the time of testing. When available, studies were also reviewed by our center's pediatric gastroenterologist (VS). The most commonly reported GI symptoms included heartburn and proximal and distal dysphagia, followed by dysmotility, weight loss, bloating, constipation, and malabsorption.

Additional clinical variables summarized included pulmonary, vascular, and musculoskeletal involvement (Table 1). Pulmonary involvement was notable, with 71% of patients affected, with more than half (57%) demonstrating abnormal pulmonary function tests (PFTs) (defined as Global Lung Initiatives values of forced vital capacity < 80% and/or diffusing capacity of the lungs for carbon monoxide < 80%), and 47% with an abnormal chest computed tomography imaging. Vascular involvement was prevalent in 96% of the cohort (mostly Raynaud phenomenon), with digital ulcers present in 67% of patients.

Medication therapy at the visit when the UCLA GIT 2.0 was completed was summarized (Table 1**)**. The most commonly prescribed disease modifying antirheumatic drug (DMARD) was mycophenolate mofetil, prescribed in 75% of patients, followed by prednisone, hydroxychloroquine, and methotrexate. Among GI drugs, proton pump inhibitors (PPIs) were the most commonly prescribed (43%), followed by histamine-2 (H2) blockers (27%). Less frequently prescribed were promotility agents, including erythromycin and metoclopramide.

Patient reported outcomes collected are also summarized in Table 1. The SHAQ-VAS are scored from 0 to 3, with higher scores indicating greater severity. The mean SHAQ-GI-VAS, which focuses on GI symptoms, had a mean score of 0.55 ± 0.79, and the mean SHAQ-DIS-VAS score was 0.99 ± 0.92.

UCLA GIT 2.0 study visit results. The mean ± SD for the UCLA GIT 2.0 scores are presented in Table 1. In this cohort with jSSc, the mean total UCLA GIT 2.0 score was 0.30 ± 0.39, which falls within the mild range. The most affected subdomain was distention and bloating (0.69 ± 0.72, considered in the moderate range), followed by the reflux subdomain (0.42 ± 0.56). All subdomains demonstrated some degree of impact, including fecal soilage and emotional wellbeing, although to a lesser extent. Consistent with findings in adult cohorts with SSc cohorts, the majority of patients with jSSc scored in the mild range. For comparison, the median total UCLA GIT 2.0 score in a large European adult cohort with SSc (n = 346 patients, 940 UCLA GIT 2.0 questionnaires) was 0.25.[9] However, in our pediatric cohort, a subset of 15.7% of patients had moderate to severe GIT total scores, ranging from 0.60 to 1.93.

The UCLA GIT 2.0 total score and its subscales showed strong correlations with the SHAQ-GI-VAS scale, with the total UCLA GIT 2.0 score having the strongest association ($r_s = 0.70$; $P < 0.001$), supporting convergent validity. There were also notable moderate correlations with the social functioning ($r_s = 0.65$; $P < 0.001$) and emotional wellbeing ($r_s = 0.62$; $P < 0.001$) subscales (Table 2). Fecal soilage ($r_s = 0.38$, $P < 0.01$) and diarrhea ($r_s = 0.29$; $P < 0.04$) subscales had weaker correlations. Additionally, the total UCLA GIT 2.0 score showed a moderate correlation

Table 1. Demographics and baseline data for 51 patients with jSSc*

Demographics and variables	Values
Age of disease onset, mean ± SD, y	9.8 ± 4.3
Disease duration at visit, mean ± SD, y	4.4 ± 3.3
Age at study visit, mean ± SD, y	14 ± 4.6
Sex, female, n (%)	36 (71)
Race and ethnicity, n (%)	
White	35 (68.6)
Black	11 (21.6)
Asian	3 (5.9)
Other	2 (3.9)
Non-Hispanic	44 (86.3)
Subtype of jSSc, n (%)	
Diffuse cutaneous	24 (47.1)
Overlap	18 (35.3)
Unclassified	5 (9.8)
Limited cutaneous	4 (7.8)
Autoantibody positivity, n (%)	
Scl-70	16 (31.4)
PM/SCL	10 (19.6)
U3 RNP	6 (11.8)
U1 RNP	5 (9.8)
Centromere	1 (2)
RNA pol III	1 (2)
RO60	1 (2)
Clinical variables	
mRSS, mean ± SD	8 ± 10.6
BMI, mean ± SD	20.98 ± 5.12
Clinical characteristics, n (%)	
Any GI involvement	39 (76.5)
Abnormal GI study[a]	37 (72.55)
Heartburn	30 (58.8)
Proximal distal dysphagia	21 (41.2)
Dysmotility	16 (31.4)
Weight loss	12 (23.5)
Bloating	11 (21.6)
Constipation	7 (13.7)
Malabsorption syndrome	3 (5.9)
Any pulmonary involvement	36 (70.6)
Abnormal PFTs[b]	29 (56.9)
Abnormal chest CT[c]	24 (47.1)
Any vascular involvement	49 (96.1)
Digital ulcer	34 (66.7)
Any musculoskeletal involvement	44 (86.3)
Myositis	22 (43.1)
Muscle weakness	22 (43.1)
UCLA GIT 2.0 scores, mean ± SD (range)	
Total	0.30 ± 0.39 (0–1.93)
Domain subsets	
Reflux	0.42 ± 0.56 (0–2.13)
Distension and bloating	0.69 ± 0.72 (0–2.75)
Fecal soilage	0.12 ± 0.38 (0–2.00)
Diarrhea	0.22 ± 0.46 (0–2.00)
Social functioning	0.25 ± 0.47 (0–2.00)
Emotional Wellbeing	0.12 ± 0.29 (0–1.22)
Constipation	0.38 ± 0.60 (0–2.50)
PROs, mean ± SD	
SHAQ-GI-VAS	0.55 ± 0.79
SHAQ-DIS-VAS	0.99 ± 0.92
Medications, n (%)	
DMARDs	
MMF	38 (75)
Prednisone	20 (39)
Hydroxychloroquine	18 (35)
Methotrexate	17 (33)
IVIG	7 (14)
Rituximab	2 (4)
Abatacept	2 (4)
Tocilizumab	2 (4)
Cyclophosphamide	1 (2)
GI agents	
PPI	22 (43)
H2 blocker	14 (27)
Erythromycin	2 (4)
Metoclopramide	1 (2)

* BMI, body mass index; CT, computed tomography; DCLO, diffusing capacity of the lungs for carbon monoxide; DMARD, disease modifying antirheumatic drug; FVC, forced vital capacity; GI, gastrointestinal; H2, histamine-2; IVIG, intravenous Ig; jSSc, juvenile systemic sclerosis; MMF, mycophenolate mofetil; mRSS, modified Rodnan Skin Score; PFT, Pulmonary Function Test; PM/SCL, polymyositis/scleroderma; PPI, proton pump inhibitor; PRO, patient-reported outcome; SHAQ-DIS-VAS, Scleroderma Health Assessment Questionnaire Disease Visual Analog Scale; SHAQ-GI-VAS, Scleroderma Health Assessment Questionnaire Gastrointestinal Visual Analog Scale; SSc, systemic sclerosis; UCLA GIT, University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract instrument.

[a] Abnormal GI studies were defined as the presence of clinically significant findings on modified barium swallow studies, gastric emptying studies, manometry, pH probe testing, or endoscopy, based on the interpretation of the pediatric gastroenterologist's or radiologist's report at the time of testing. When available, studies were also reviewed by our center's pediatric gastroenterologist (VS).

[b] Global Lung Initiatives values of FVC < 80% and/or DL_{CO} < 80%; if multiple PFTs were available, the one performed closest to the time of the UCLA GIT 2.0 study visit was recorded.

[c] The presence of ground glass opacities, reticulations, or honeycombing consistent with interstitial lung disease; if multiple chest CTs were available, the one performed closest to the time of the UCLA GIT 2.0 study visit was recorded.

with the overall scleroderma disease assessment PRO, the SHAQ-DIS-VAS (r_s = 0.60; P < 0.001).

UCLA GIT 2.0 longitudinal results. From the group, 22 patients had a follow-up UCLA GIT 2.0 assessment and corresponding SHAQ-VAS completed on an average of 13 ± 3.5 months. Scores from all subdomains decreased at follow-up, except for emotional wellbeing (Table 3). These were expressed in mean values to reflect the change in the score and the minimally clinically important difference (MCID) as developed by Khanna et al.[14] Mean UCLA GIT 2.0 scores significantly improved from 0.34 ± 0.45 (95% confidence interval [CI], 0.14–0.54) to 0.23 ± 0.341 (95% CI, 0.07–0.38; P = 0.041, paired t-test). The mean change was –0.11 (95% CI, –0.214 to –0.011), with a moderate effect size (Cohen's d, –0.47). The normality of within-participant differences was assessed using Q–Q plots and the Shapiro–Wilk test (W = 0.94; P = 0.18), both of which supported the assumption of normality and justified the use of paired t-test (parametric) analyses to compare GIT scores between time points. However, the change did not meet the

Table 2. Convergent construct validity of the UCLA GIT 2.0 instrument*

SHAQ-VAS (n = 51)	Spearman's (r_s)	Correlation coefficient scale	95% CI	*P* value
SHAQ-GI-VAS				
UCLA GIT total	0.70	Strong	0.493–0.827	<0.001
Domain subscale				
Reflux	0.56	Moderate	0.310–0.733	<0.001
Distension and bloating	0.59	Moderate	0.357–0.759	<0.001
Fecal soilage	0.38	Weak	0.108–0.605	0.01
Diarrhea	0.29	Weak	0.003–0.527	0.04
Social functioning	0.65	Moderate	0.427–0.795	<0.001
Emotional wellbeing	0.62	Moderate	0.392–0.777	<0.001
Constipation	0.42	Moderate	0.150–0.634	<0.001
SHAQ-DIS-VAS				
UCLA GIT total	0.60	Moderate	0.367–0.764	<0.001

* Comparisons with the SHAQ-GI-VAS and SHAQ-DIS-VAS used Spearman's correlation. The correlation coefficient value scale was reported as weak (< 0.4), moderate (0.4–0.7), or strong (> 0.7). CI, confidence interval; SHAQ-DIS-VAS, Scleroderma Health Assessment Questionnaire Disease Visual Analog Scale; SHAQ-GI-VAS, Scleroderma Health Assessment Questionnaire Gastrointestinal Visual Analog Scale; UCLA GIT, University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract instrument.

established MCID for improvement (–0.20). Distention and bloating was the only subdomain that achieved both statistical and MCID threshold significance, with the social functioning and constipation subdomains achieving a change meeting the MCID threshold.

The overall disease status in this subgroup of patients improved over time reflected by the significant decrease in the accepted surrogate for disease status in patients with SSc, the mRSS, reflecting improvement in disease status with the mean score dropping from 8.18 at first center visit to 3.5 at follow-up ($P < 0.001$), and also meeting the established MCID threshold for improvement of the mRSS was –2.86, and this criterion was met.[15] Outcomes of GI status, such as BMI and the SHAQ-GI-VAS, showed improvement with decreasing scores respectively.

DISCUSSION

This is the first study to apply the UCLA GIT 2.0 instrument, originally validated in adults, to a cohort of patients with jSSc. Our primary aim was to evaluate its use for patients with jSSc, including face validity, convergent construct validity and preliminary responsiveness. Given the high prevalence of GI involvement in patients with SSc, understanding its impact on pediatric patients is essential for improving care and monitoring disease

Table 3. Longitudinal changes in PROs and clinical variables in patients with jSSc over one year*

Outcome Measure	Baseline, mean score	Follow-up, mean score	Difference	Cited MCID for improvement	Does it meet MCID threshold?	*P* value
UCLA GIT 2.0						
Total	0.34	0.23	−0.11	−0.20	NO	**0.039**
Domain subscales						
Reflux	0.41	0.34	−0.07	−0.26	NO	0.441
Distension and bloating	0.83	0.53	−0.30	−0.14	YES	**0.049**
Fecal soilage	0.23	0.09	−0.14	−0.19	NO	0.083
Diarrhea	0.18	0.14	−0.04	−0.19	NO	0.665
Social functioning	0.26	0.12	−0.14	−0.07	YES	0.062
Emotional wellbeing	0.13	0.13	0.00	−0.36	NO	0.940
Constipation	0.42	0.25	−0.17	−0.17	YES	0.101
SSc PROs						
SHAQ-GI-VAS	0.88	0.62	−0.26	–	–	0.131
SHAQ-DIS-VAS	0.46	0.57	0.11	–	–	0.498
Clinical outcomes						
mRSS	8.18	3.50	−4.68	−2.86	YES	**0.001**
BMI	20.37	21.00	0.63	–	–	0.077

* Comparison of baseline and one year follow-up data in 22 paired patients with jSSc, including PROs and clinical measures. MCID thresholds applied when available, with statistical significance determined by paired *t*-tests indicated by bold values in the table ($P < 0.05$). BMI, body mass index; jSSc, juvenile systemic sclerosis; MCID, minimally clinically important difference; mRSS, modified Rodnan Skin Score; PRO, patient-reported outcome; SSc, systemic sclerosis; UCLA GIT, University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract instrument.

burden. We analyzed cross-sectional data from 51 patients with jSSc and observed 22 patients with jSSc longitudinally to assess changes in GI symptoms and related outcomes.

Upper GI tract symptoms, particularly reflux and distention and bloating, were reported as the most affected, aligning with the clinical characteristics observed in adult cohort (eg, heartburn, dysphagia, and dysmotility). This symptom pattern supports the face validity of the UCLA GIT 2.0 in patients with jSSc. The highest subdomain impact was in distention and bloating, followed by reflux, mirroring findings from cohorts with adult SSc across North America, Europe, and Australia, although absolute scores were slightly lower in patients with jSSc.[9–11,14] In contrast, lower GI involvement (eg, fecal soilage, constipation) was less prominent in our cohort, which may reflect age-related differences in disease manifestation or patient-reported symptom burden.

Despite the high prevalence of upper GI symptoms, only 27% of patients were prescribed H2 blockers, and 43% were prescribed PPIs, lower than expected based on symptom burden. This suggests a potential gap in symptom recognition and treatment. This gap is clinically important because upper GI dysfunction has been linked to interstitial lung disease in patients with jSSc,[5] which is prevalent in this cohort. Pediatric patients may underreport GI symptoms due to difficulty verbalizing symptoms or embarrassment. Administering the UCLA GIT 2.0 as a PRO at the start of the clinic visit may help uncover unreported symptoms and facilitate earlier intervention.

The UCLA GIT 2.0 demonstrated strong convergent validity in patients with jSSc, with significant correlation between total UCLA GIT 2.0 scores and the SHAQ-GI-VAS (r_s = 0.70, P < 0.001), consistent with adult SSc validation studies[11] because the SHAQ-GI-VAS reflects the total impact of GI issues interfering with daily function. Moderate correlations were observed with social functioning (r_s = 0.65) and emotional wellbeing (rs = 0.62) subdomains, further reinforcing its relevance in assessing QoL impact, and aligning with findings from a large Australian cohort with SSc (n = 907).[10] Additionally, the total UCLA GIT 2.0 score correlated with the SHAQ-DIS-VAS, supporting the broader role of GI burden in overall patient-perceived disease severity, as demonstrated in the North American CARRA cohort with jSSc.[1]

Among the 22 patients with longitudinal data, overall clinical improvement was reflected in reduced mRSS and SHAQ-GI-VAS scores after a mean follow-up of 12 months. The UCLA GIT 2.0 score demonstrated statistical improvement over time; however, most domains did not reach published adult-derived MCID thresholds. Importantly, the distension and bloating subdomain, the most impacted at baseline study visit in both the cohort with jSSc and comparable adult cohorts with SSc, demonstrated both statistically and clinical meaningful improvement over time, supporting preliminary responsiveness of the UCLA GIT 2.0 in patients with jSSc. Although other subdomains of the UCLA GIT 2.0 also showed improvement in patients with jSSc, these changes were not statistically significant, likely due to the small sample size, a limitation of this portion of the study.

This study has several notable strengths. It represents the first application of a validated GI PRO (UCLA GIT 2.0) in a pediatric cohort with SSc, demonstrating its potential relevance in juvenile-onset disease. The use of a prospective, registry-based design enabled standardized data collection across a range of disease durations, and the inclusion of complementary outcome measures, such as the SHAQ-VAS, allowed for a robust assessment of convergent construct validity. The UCLA GIT 2.0 also proved feasible in a clinical setting, with most patients completing the questionnaire within five to seven minutes following brief guidance, based on over a decade of routine use at our center.

However, important limitations should be acknowledged. The study was conducted at a single tertiary center and included a relatively modest sample size, which may limit generalizability. In addition, the timing of the UCLA GIT 2.0 survey administration varied among patients; for some, it reflected early disease, whereas for others, it was completed after substantial treatment exposure, potentially affecting symptom reporting. Another limitation is the application of adult-derived MCID thresholds to pediatric patients because the relevance of these cutoffs in children is uncertain, and meaningful symptom change may differ between pediatric and adult populations.

Future directions include pediatric-specific validation of the UCLA GIT 2.0 through qualitative interviews and cognitive debriefing to assess face and content validity. Additionally, efforts to define pediatric-specific MCID thresholds using global patient-reported change anchors (such as the five-point Likert scale used in the SSc validation studies among adults) would be helpful. Broader multicenter efforts to evaluate performance and sensitivity to change over time in larger cohorts with jSSc are also warranted.

This study provides the first evidence supporting the utility of the UCLA GIT 2.0 in patients with jSSc, demonstrating similar performance in patients with adult SSc with consistent subscale relevance, and strong correlations with established GI QoL measures. These findings support the face and convergent construct validity of this tool in capturing GI impact in patients with jSSc. Clinically, the most prevalent symptoms include esophageal dysfunction with associated gastroesophageal reflux, delayed gastric emptying, and small bowel dysfunction, as reflected in the most affected subdomains of the UCLA GIT 2.0, reflux and distension or bloating. Preliminary evidence of sensitivity to change especially in the distension and bloating domain supports further longitudinal validation. Future efforts should focus on refining this tool for pediatric use and integrating it into routine clinical practice to enhance symptom detection, patient communication, and individualized treatment planning.

ACKNOWLEDGMENTS

We thank all National Registry for Childhood Onset Scleroderma enrollees for participation in this study.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Torok 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. Stevens BE, Torok KS, Li SC, et al; Childhood Arthritis and Rheumatology Research Alliance Registry Investigators. Clinical characteristics and factors associated with disability and impaired quality of life in children with juvenile systemic sclerosis: results from the Childhood Arthritis and Rheumatology Research Alliance legacy registry. Arthritis Care Res (Hoboken) 2018;70(12):1806–1813.
2. McMahan ZH, Kulkarni S, Chen J, et al. Systemic sclerosis gastrointestinal dysmotility: risk factors, pathophysiology, diagnosis and management. Nat Rev Rheumatol 2023;19(3):166–181.
3. Foeldvari I, Klotsche J, Kasapcopur O, et al. Differences sustained between diffuse and limited forms of juvenile systemic sclerosis in an expanded international cohort. Arthritis Care Res (Hoboken) 2022; 74(10):1575–1584.
4. Scalapino K, Arkachaisri T, Lucas M, et al. Childhood onset systemic sclerosis: classification, clinical and serologic features, and survival in comparison with adult onset disease. J Rheumatol 2006;33(5): 1004–1013.
5. Ambartsumyan L, Zheng HB, Iyer RS, et al. Relationship between esophageal abnormalities on fluoroscopic esophagram and pulmonary function testing in juvenile systemic sclerosis. Arthritis Care Res (Hoboken) 2019;71(11):1444–1449.
6. Foeldvari I, Torok KS. Review for best practice in clinical rheumatology juvenile systemic sclerosis - updates and practice points. Best Pract Res Clin Rheumatol 2021;35(3):101688.
7. Spiegel BM. Patient-reported outcomes in gastroenterology: clinical and research applications. J Neurogastroenterol Motil 2013;19(2): 137–148.
8. Khanna D, Hays RD, Maranian P, et al. Reliability and validity of the University of California, Los Angeles Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. Arthritis Rheum 2009; 61(9):1257–1263.
9. Zampatti N, Garaiman A, Jordan S, et al. Performance of the UCLA Scleroderma Clinical Trials Consortium Gastrointestinal Tract 2.0 instrument as a clinical decision aid in the routine clinical care of patients with systemic sclerosis. Arthritis Res Ther 2021;23(1):125.
10. Quinlivan A, Hansen D, Stevens W, et al. Prevalence and outcomes of gastrointestinal manifestations in an Australian scleroderma cohort. Arthritis Care Res (Hoboken). 2024;76(12):1686–1695.
11. Baron M, Hudson M, Steele R, et al. Canadian Scleroderma Research Group. Validation of the UCLA Scleroderma Clinica Trial Gastrointestinal Tract Instrument version 2.0 for systemic sclerosis. J Rheumatol. 2011;38(9):1925–1930.
12. Van den Hoogen F, Khanna D, Fransen J, et al. 2013. Classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–2747.
13. Steen VD, Medsger TA Jr. The value of the health assessment questionnaire and special patient-generated scales to demonstrate change in systemic sclerosis patients over time. Arthritis Rheum 1997;40(11):1984–1991.
14. Khanna D, Furst DE, Maranian P, et al. Minimally important differences of the UCLA Scleroderma Clinical Trial Consortium Gastrointestinal Tract Instrument. J Rheumatol 2011;38(9):1920–1924.
15. Khanna D, Clements PJ, Volkmann ER, et al. Minimal Clinically important differences for the modified Rodnan Skin Score: results from the Scleroderma Lung Studies (SLS-I and SLS-II). Arthritis Res Ther 2019;21(1):23.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 456–468
DOI 10.1002/acr.25619

American College of Rheumatology
Empowering Rheumatology Professionals

Monocyte LOXHD1 and RHOB Expression Predictive of Progressive Systemic Sclerosis–Associated Interstitial Lung Disease

Cristina M. Padilla,[1] Robert Lafyatis,[1] Kevin F. Gibson,[1] Christina Morse,[1] Eleanor Valenzi,[1] Xiaoping Chen,[1] Xiaoyun Li,[1] Yanwu Zhao,[1] Yongseok Park,[1] Dana Ascherman,[1] Jonathan Minden,[2] Robyn Domsic,[1] Yingze Zhang,[1] and Daniel J. Kass[1]

Objective. A leading cause of death among patients with scleroderma (SSc), interstitial lung disease (ILD) remains challenging to prognosticate. The discovery of biomarkers that accurately determine which patients would benefit from close monitoring and aggressive therapy would be an essential clinical tool. We aimed to identify gene signatures that would predict progressive ILD.

Methods. We compared previously identified serum biomarkers, bulk peripheral blood mononuclear cell (PBMC) RNA gene expression (39 patients with progressive SSc-ILD and 43 patients with stable SSc-ILD), and single-cell RNA gene expression of PBMC (13 patients with progressive SSc-ILD, 14 patients with stable SSc-ILD, and 6 control participants).

Results. In previous studies, male sex, Krebs von den Lungen 6 levels, and C-reactive protein levels were predictors of progressive disease in patients. Monocyte expression of lipoxygenase homology domains 1 (LOXHD1) and Ras homolog gene family, member B (RHOB) strongly predicted progression, suggesting a key role in immune dysregulation. *LOXHD1* and related genes were enriched in inflammatory gene networks, which may support monocyte-associated inflammation in patients. RHOB was consistently up-regulated in both CD14+ and CD16+ monocytes across single-cell and bulk RNA data, underscoring robust association with progressive disease. In contrast, ataxin-2-like (ATXN2L) protein was down-regulated in patients with progressive disease, suggesting dysregulation of cellular stress responses. Additionally, S100 calcium-binding protein A12 (S100A12), chitinase-3-like protein 1 (CHI3L1), and matrix metalloproteinase-9 (MMP9), previously linked to tissue remodeling, were again associated with progression in bulk RNA sequencing and previous microarray studies, reinforcing their role in fibrosis.

Conclusion. With several potential biomarkers of progressive disease in patients, the next step includes validation in larger, multicenter cohorts for use in clinical decision-making.

INTRODUCTION

Interstitial lung disease (ILD) in patients with systemic sclerosis (SSc) is the most lethal complication, occurring in patients with both limited cutaneous and diffuse cutaneous disease. The course of SSc-ILD is quite variable. In some patients, the disease is mild and not progressive. Patients with mild disease on presentation have a better prognosis. On the other hand, patients with anti-Scl-70 autoantibodies tend to have more severe, progressive disease.

Several serum biomarkers of disease severity have been explored. C-reactive protein (CRP) is the most widely available and best studied prognostic biomarker. We have recently reviewed other biomarkers: surfactant protein-D (SP-D), Krebs

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, the NIH (grant 1P50-AR-080612), and by Mitsubishi Inc. Dr Padilla's work is supported by the Physician-Scientist Institutional Award held by the University of Pittsburgh funded by the Burroughs Wellcome Fund.

[1]Cristina M. Padilla, MD, Robert Lafyatis, MD, Kevin F. Gibson, MD, Christina Morse, MBA, Eleanor Valenzi, MD, Xiaoping Chen, MS, Xiaoyun Li, MD, Yanwu Zhao, BS, Yongseok Park, PhD, Dana Ascherman, MD, Robyn Domsic, MD, Yingze Zhang, PhD, Daniel J. Kass, MD: University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Jonathan Minden, PhD: Carnegie Mellon University, Pittsburgh, Pennsylvania.

Drs Padilla and Lafyatis are co-first authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25619).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25619.

Address correspondence via email to Cristina M. Padilla, MD, at padillac@pitt.edu.

Submitted for publication February 14, 2025; accepted in revised form July 22, 2025.

SIGNIFICANCE & INNOVATIONS

- Up-regulated expression of lipoxygenase homology domains 1 and Ras homolog gene family, member B and down-regulated expression of ataxin-2-like protein in monocytes are associated with progressive systemic sclerosis (SSc) interstitial lung disease (ILD) in patients and could be prognostic biomarkers.
- S100 calcium-binding protein A12, chitinase-3-like protein 1, and matrix metalloproteinase-9 are part of a cluster associated with progressive SSc-ILD in patients and could represent true biomarkers of SSc-ILD.
- Biomarkers that track using disease activity or prognostication of SSc-ILD in patients are invaluable clinical and research tools, which could lead to improved outcomes and effective therapies.

von den Lungen 6 (KL-6), osteopontin (SPP1), and C-C motif chemokine 18 (CCL18), which have also been studied as serum biomarkers; however, none are strongly predictive of a decline in pulmonary function.[1] Thus, the discovery of prognostic biomarkers of SSc-ILD would provide a significant new tool for clinicians to decide which patients to treat aggressively and for clinical trialists to help stratify patients for inclusion into clinical trials.

Multiple lines of evidence support the importance of autoimmunity and inflammation in patients with SSc and SSc-ILD. We have shown previously that patients with SSc-ILD have profound changes in the inflammatory cell populations in the lungs.[2] This includes marked alterations in gene expression by three different monocyte-macrophage populations. Alveolar macrophages are marked by *FABP4*, *INHBA*, and *SERPING1*, and monocyte-like macrophages marked by *CD14*, *FCN1*, *IL1B*, and *IL1R2*.[2] A third population of macrophages (SPP1 macrophages) show markedly up-regulated expression of *SPP1*, *MMP9*, and *CCL18* in patients with SSc-ILD as well as in patients with idiopathic pulmonary fibrosis (IPF).[2–4] In view of the observation that bleomycin-induced fibrosis in SPP1-deleted mice is attenuated and a similar macrophage phenotype is seen in other fibrotic diseases, these macrophages appear profibrotic.[5] T and natural killer (NK) cell populations in patients with SSc-ILD also show altered gene expression compared with lungs from healthy controls (HC).[6]

Because lung biopsies and bronchoalveolar lavage are not routinely required diagnostically in patients with SSc-ILD, gene expression by peripheral blood mononuclear cells (PBMCs) provides a potential surrogate for assessing immune cell function and for discovery of prognostic biomarkers in patients with SSc-ILD. Analyzing PBMC gene expression by microarray, we previously reported that some PBMCs from patients with SSc show an interferon (IFN) signature.[7,8] We have also previously identified genes up-regulated in PBMCs from patients with SSc with new onset isolated pulmonary arterial hypertension (*CCR1*, *IL13RA1*, *TIMP2*, *AIF1*, *ICAM1*, and *IFNGR1*).[8] Here, we compare previously identified serum biomarkers, bulk PBMC RNA gene expression, and single-cell RNA gene expression in patients with progressive SSc-ILD versus patients with stable SSc-ILD. These studies highlight several gene expression prognostic biomarkers, particularly monocyte expression of *LOXHD1* and *RHOB*, for patients with progressive SSc-ILD.

PATIENTS AND METHODS

Study cohort. Blood samples were obtained from patients with SSc in the Dorothy P and Richard P Simmons Center for Interstitial Lung Disease at the University of Pittsburgh Medical Center (UPMC; the Simmons Center) or the Falk Rheumatology Clinic, who completed baseline evaluations and consented to experimental studies using their venous phlebotomy samples. Complete descriptions of recruitment and clinical evaluations were previously detailed.[9,10] Spirometry and diffusing capacity were performed using American Thoracic Society standards and standard reference equations.[11–13] This study was approved by the Institutional Review Board for Human Subject Research at the University of Pittsburgh (STUDY20030223). All participants provided written informed consent (in accordance with the Declaration of Helsinki) before their participation in this study.

Archival PBMC samples and matched to date sera from the Simmons Center, and prospectively collected blood samples from patients seen in the clinic were studied. Samples collected at baseline (enrollment blood collection) from patients with SSc-ILD were stratified into progressive disease or stable disease based on one of five criteria: (1) death within two years of enrollment from any cause; (2) lung transplant within two years after enrollment; (3) forced vital capacity (FVC) less than 50% at time of or within two years after enrollment; (4) greater than 10% decline in FVC over two years comparing both before and after blood collection; or (5) initiated on new immunosuppressive medication for ILD within three months of blood collection. Patients not meeting one of these criteria were considered to have stable disease at the time of enrollment. Patients for whom criteria were ambiguous were adjudicated as stable or progressive by chart review by Drs. Lafyatis and Kass. Demographic information and autoantibody reactivities were recorded for each patient when available from the medical chart. Figure 1A shows how both prospective and archival samples were used.

Serum biomarker analyses. CRP, KL-6, and SP-D levels were analyzed using citrate plasma collected at the same time as PBMC from each patient. CRP and SP-D were analyzed using Quantikine enzyme-linked immunosorbent assay kits (DCRP00 and DSFPD0, R&D Systems). KL-6 levels were analyzed using Luminex kit (LXSAHM-01, R&D Systems) and Bioplex 200 (Bio-Rad Labs). Sera collected at the same time as PBMC from each

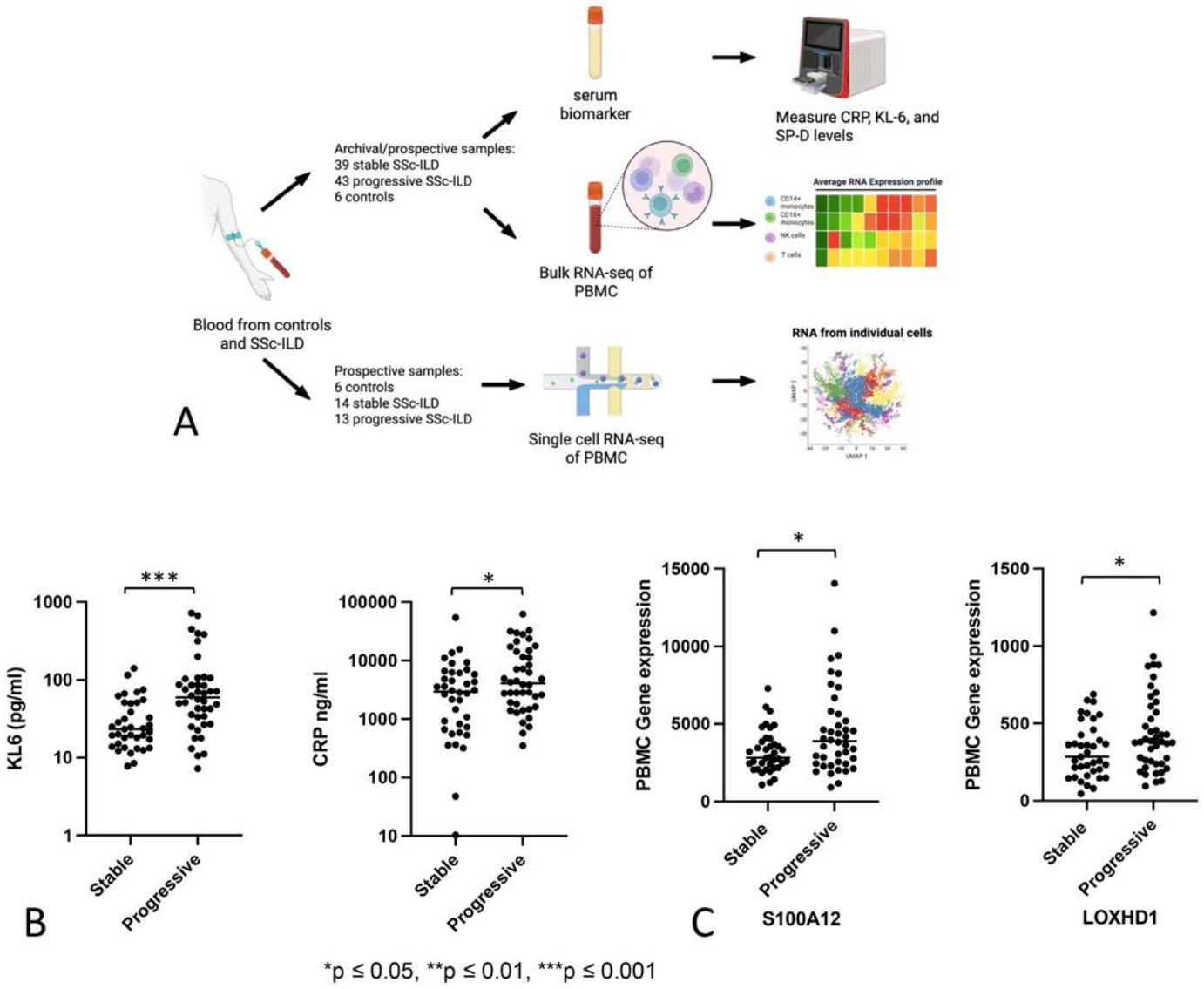

Figure 1. (A) Schematic overview of sample processing and data generation. (B) Bar plots comparing the serum levels of KL-6 ($P = 0.0004$) and CRP ($P = 0.03$) between patients with stable SSc-ILD and patients with progressive SSc-ILD using two-sided Mann-Whitney. The horizontal line represents the average gene expression. (C) Bar plots exhibiting scRNA-seq gene expression of monocyte-like myeloid cells from patients with stable SSc-ILD and progressive SSc-ILD. These genes were found to have increased gene expression compared with healthy control (Supplementary Figure 1). *$P \leq 0.05$. **$P \leq 0.01$. ***$P \leq 0.001$. CRP, C-reactive protein; KL-6, Krebs von den Lungen 6; NK, natural killer; PBMC, peripheral blood mononuclear cells; scRNA-seq, single-cell RNA sequencing; SP-D, surfactant protein-D; SSc-ILD, scleroderma with interstitial lung disease. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25619/abstract.

participant were used for analyzing creatine phosphokinase (CPK) levels through the UPMC central laboratory using an International Federation of Clinical Chemistry and Laboratory Medicine – creatine kinase (*N*-acetylcysteine) kit on Beckman Coulter AU680 and AU5800 analyzers. The two-sided Mann-Whitney U test determined statistical significance.

RNA isolation. For the historical samples, PBMCs were isolated using citrate cell preparation tube (CPT) vacutainer tubes (Beckman-Dickerson) and stored in either Trizol or RNAlater (Thermo-Fisher) at –80°C until RNA isolation. Samples stored in RNAlater were lysed in Trizol during RNA isolation procedure. RNA isolation was followed standard protocol using Qiagen Rneasy mini kit (Qiagen). All RNA samples were quantified using Agilent RNA 6000 Nano Kit and Bioanalyzer (Agilent 2100). All samples passed quality control with an integrity number greater than eight.

RNA sequencing library generation and sequencing. RNA was assessed for quality (Agilent TapeStation 4150) and RNA concentration quantified (Qubit FLEX fluorometer). Libraries were generated with the Illumina Stranded messenger RNA Library Prep kit (catalog #20040534) according to the manufacturer's instructions. Briefly, 100 ng of input RNA was used for each sample. Following adapter ligation, 13 cycles of indexing

PCR were completed, using Integrated DNA Technologies (IDT) for Illumina RNA UD Indexes (Illumina, catalog #20040553-6). Libraries were quality and quantity and assessed (Qubit FLEX fluorometer and Agilent TapeStation 4150). Libraries were normalized and pooled to 10 nM by calculating the concentration based off the fragment size (base pairs) and the concentration (ng/μl) of the libraries. Sequencing was performed on an Illumina NovaSeq 6000 (UPMC Genome Center). Libraries were sequenced on an S2 flow cell with a target of ~50 million reads per sample.

Analysis of bulk RNA-seq data. PBMC RNA gene expression was analyzed by RNA sequencing (RNA-seq). We then used three approaches to examine the RNA-seq data: Examining genes that were differentially expressed between stable and progressive patients using DESeq2, examining pathways regulated by these genes with uncorrected $P < 0.05$ using Gene Ontology (GO). False discovery rate (FDR) <5% was considered statistically significant. Genes were then clustered by meeting an expression threshold for detection using Cluster 3.0 and Java TreeView. The uncorrected *P* value was used to identify potential candidate genes.

Single-cell RNA sequencing of PBMC. PBMC were collected from patients and HCs and stored frozen in 10% dimethyl sulfoxide. Frozen PBMC were thawed in batches of 10 samples, including two from HCs and three or four each of samples from patients with progressive SSc-ILD and stable SSc-ILD. Samples were labeled with barcoded antibodies for multiplexing samples (10X Genomics),[14] as described previously.[15] Targeting 5,000 cells sample, cells were partitioned, and single-cell libraries prepared using the V3 3′ chemistry (10X Genomics), according to the 10X Genomics protocol. Libraries were sequenced, and sequences demultiplexed and aligned in Cell Ranger (10X Genomics).

Analysis of single-cell RNA-seq data. Cell gene expression matrixes were generated from the aligned sequence data. The Model-based Analysis of Single-cell Transcriptomics (MAST) statistical algorithm feature with adjusted $P < 0.05$ in Seurat version 4 was used to discover differentially expressed genes in the single-cell RNA-seq (scRNA-seq) data comparing cells from patients with progressive SSc-ILD with cells from patients with stable SSc-ILD cluster by cluster. Cells were plotted by transcriptome using the uniform manifold approximation and projection (UMAP) algorithm for dimensionality reduction, annotated using Azimuth (Figure 3A),[15] and confirmed cluster identities using known marker genes (Supplementary Figure 8). We used Kruskal-Wallis test to compare samples from control, stable SSc-ILD, and progressive SSc-ILD. Dunn's test was performed following Kruskal-Wallis to identify which specific groups differ.

RESULTS

Male sex and shorter disease duration associated with progressive SSc-ILD. Ages were well-balanced between patients with progressive SSc-ILD (56.59 years) and patients with stable SSc-ILD (58.79 years). Progressive SSc-ILD was strongly skewed toward patients who were men and with shorter disease duration (Table 1). Men with progressive disease represented 46.5% of patients with progressive SSc-ILD but only 15.4% of patients with stable SSc-ILD ($P = 0.0038$). Disease duration was significantly shorter in patients with progressive disease (8.86 years) compared with patients with stable disease (15.55 years; $P = 0.01242$). As would be expected from the criteria for progressive disease, the baseline FVC was significantly lower in patients with progressive SSc-ILD (74.1% predicted) compared with patients with stable SSc-ILD (60.7% predicted).

Elevated CRP and KL-6 levels in patients with progressive SSc-ILD. We examined the sera to see if KL-6, CRP, SP-D, or CPK levels were predictive for progressive SSc-ILD. Both KL-6 and CRP levels were statistically significantly higher in patients with progressive SSc-ILD compared with patients with stable SSc-ILD (Table 1, Figure 1B). Notably, there was little correlation between CRP and KL-6 levels ($r^2 = 0.0178$; $P > 0.05$), suggesting that these two biomarkers provide somewhat complementary information regarding prognosis. SP-D levels failed to show a difference between patients with progressive SSc-ILD and patients with stable SSc-ILD; however, many of the SP-D measures were at the limit of detection, perhaps explaining why these results are not consistent with some past studies. CPK, found in an early analysis of the GENOSIS cohort to be protective for the presence of SSc-ILD[9] and was later shown to be weakly associated with a lower baseline FVC after adjustment for follow up time,[10] did not show a significant difference between patients with progressive SSc-ILD and patients with stable SSc-ILD in our cohort.

In addition, we looked at autoantibodies in the sera of these patients, detectable in 65 of 80 patients. Patients with anti-Scl-70 autoantibodies had a strong trend toward having progressive disease ($P = 0.0783$), whereas patients with anti-centromere (ACA), anti-U1 or anti-U2, and anti-Th or anti-To antibodies were associated with less progressive disease ($P = 0.1796$ for each; Table 1). Other autoantibody specificities did not associate with patients with stable SSc-ILD or patients with progressive SSc-ILD. These autoantibody and serum data indicated that our designation of patients with stable SSc-ILD and patients with progressive SSc-ILD were consistent with previous studies of patients with SSc-ILD.

Monocyte-like cells increased expression of lipoxygenase homology domains 1. We examined the most differentially expressed genes (DEGs) between patients with progressive SSc-ILD and patients with stable SSc-ILD and ranked these by uncorrected $P < 0.05$ (Supplementary Table 1). Although

Table 1. Patient demographics, clinical features, serum biomarkers, and autoantibodies*

	Stable (n = 39)	Progressive (n = 43)	Nominal *P*-value[a]
Sex, male	6	20	**0.0038**
Age, y	58.79	56.59	0.4777
Disease duration,[b] y	15.55	8.86	**0.01242**
Limited/diffuse[c]	26/9	22/12	0.4251
Baseline predicted FVC, %	70.1	60.7	**<0.0001**
MRSS[d]	4.62	6.03	0.08726
CRP	5053.7	9663.3	**0.0394**
KL-6	36.0	119.1	**0.00044**
SP-D	11.0	13.8	0.4965
CPK	66.2	70.8	0.88866
Autoantibodies[e]			
anti-Scl-70	6	14	0.0783
anti-Th/To	7	3	0.1796
anti-RNAPol3	5	4	0.7292
anti-centromere	7	3	0.1796
anti-Ku	1	1	1.000
anti-PM/Scl	3	4	1.000
anti-U3	2	1	0.6018
anti-U1/U2 RNP	4	1	0.1842
anti-Ro	1	4	0.3622

The bolded values are the *p*-values that reached statistical significance ($p < 0.05$).
* One patient showed autoantibodies to both Ku and RNApol3. Bold *P* values indicate statistical significance. CPK, creatine phosphokinase; CRP, C-reactive protein; FVC, forced vital capacity; KL-6, Krebs von den Lungen 6; MRSS, modified Rodnan Skin Score; RNP, ribonucleoprotein; SP-D, surfactant protein-D.
[a] Two-sided Mann-Whitney, or two-sided Fisher's exact test for binomial values.
[b] Patients in stable disease group, n = 31; patients in progressive disease group, n = 32.
[c] Patients in stable disease group, n = 35; patients in progressive disease group, n = 34.
[d] Patients in stable disease group, n = 32; patients in progressive disease group, n = 32.
[e] For all autoantibodies measures, patients in stable disease group (n = 38), patients in progressive disease group (n = 42).

none of these genes reached a corrected *P* value or FDR of less than five percent, several genes appeared potentially of interest as biomarkers based on their increased expression in SSc-ILD monocyte-like myeloid cells in our previously reported single-cell RNA sequencing (scRNA-seq) data. Among the genes elevated in patients with progressive SSc-ILD, *S100A12* stood out for its previously reported role in patients with progressive IPF,[11] highlighting its potential as a shared biomarker of fibrotic lung disease. Similarly, *S100A9* and *FPR2* were up-regulated, associated with immune activation, which may suggest a proinflammatory state contributing to disease progression (Supplementary Figure 1). Notably, *LOXHD1*, which also showed upregulation in our scRNA-seq data set (Figure 1C, Supplementary Figure 1), may also reflect proinflammatory programming relevant to fibrotic progression. Among genes with nominal significance (uncorrected $P < 0.05$), *MMP9* emerged as one of the most up-regulated in patients with progressive SSc-ILD versus patients with stable SSc-ILD, showing an 8.76-fold increase (Supplementary Figure 2, Supplementary Table 1). Matrix metalloproteinase-9 (MMP9) is a known biomarker for SSc-ILD and may reflect ongoing tissue remodeling and fibrotic activity.[12] We have previously described markedly up-regulated MMP9 by SPP1 macrophages,[3,4] a macrophage subset associated with increased SPP1 expression and fibrosis, in lungs from patients with SSc-ILD (Supplementary Figure 2). SPP1 also appeared elevated in the PBMCs from patients with progressive SSc-ILD, although its expression level was lower than the threshold that we employed to filter noise in gene expression. *SPP1* is a gene known to be up-regulated by profibrotic macrophages in patients with IPF and patients with SSc-ILD.[3,4] Ras homolog gene family, member B (RHOB) was also seen up-regulated.

We evaluated whether genes with nominal significance were enriched in specific biologic pathways. GO indicated several statistically significant pathways (Supplementary Table 2). Several enriched pathways were inflammation-related, including inflammatory response (6.13-fold), chronic inflammatory response (65.37-fold), respiratory burst (44-fold), response to lipopolysaccharide (LPS) (7.06-fold), and neutrophil chemotaxis (17.16-fold). Key genes involved included *CCL4L2*, *F12*, *FPR2*, *S100A8*, *FOS*, *LYN*, *S100A9*, *CXCL1*, *VNN1*, and *S100A12*.

We also examined down-regulated genes (Supplementary Table 3). Again, none of these genes reached an FDR of less than five percent. However, several T cell receptor genes and the *IL10RA* gene showed decreased expression in T cells and monocyte-like macrophages, respectively, in lungs from patients with SSc-ILD (Supplementary Figure 3). Pathway analysis of down-regulated genes indicated pathways centered around nucleic acid and messenger RNA metabolism (Supplementary Table 3).

Inflammatory gene cluster more highly expressed by PBMCs from patients with progressive disease. Clustering of genes revealed several groups of highly correlated genes. Most prominently, IFN-regulated genes were expressed

by a subset of patients, but none of these genes were up-regulated in patients with progressive SSc-ILD (Figure 2, cluster 1). A second cluster contained *S100A8*, *S100A9*, *S100A12*, and *FPR2*, genes up-regulated in patients with progressive disease and highlighted in the GO pathway analysis above (Figure 2, cluster 2). The up-regulated gene most highly correlated with patients with progressive disease was *RP11-153M7* (a *TLR2*-related pseudogene), clustered with other genes, but none of the other genes in this cluster distinguished progressive disease (Figure 2, cluster 3). A fourth cluster included *MMP9*, *BMX*, *CXCR1*, *CXCR2*, and *FCGR3B/CD16B* genes. These genes were both more highly expressed or trended strongly toward higher expression in PBMCs from patients with progressive SSc-ILD. A fifth cluster included *LOXHD1* and *TUBB2A*, both more highly expressed in PBMCs from patients with progressive SSc-ILD (Supplementary Figure 4 for full gene list).

Several additional clusters of genes showed higher expression in patients with progressive SSc-ILD. A cluster of Y chromosome genes were easily recognized as only expressed by male patients, and a cluster of hemoglobin and other genes associated with red blood cells appeared to be coming from reticulocytes (Supplementary Figure 5), possibly related to hypoxemia or increased bleeding in these patients often known to have gastric antral vascular telangiectasia. In addition, a cluster included the

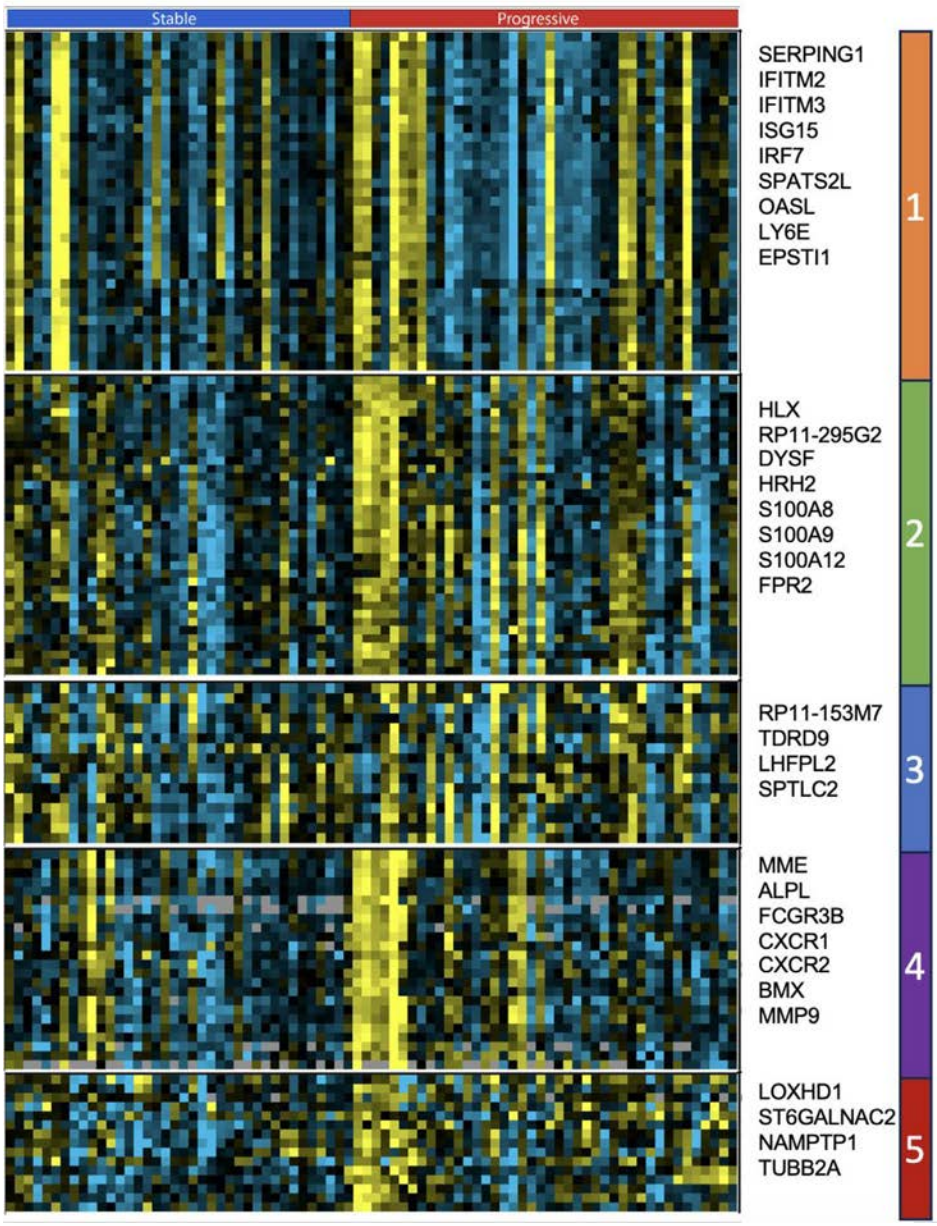

Figure 2. Gene cluster heatmap showing different clusters according to their inflammatory gene signature from PBMCs. PBMC, peripheral blood mononuclear cells. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25619/abstract.

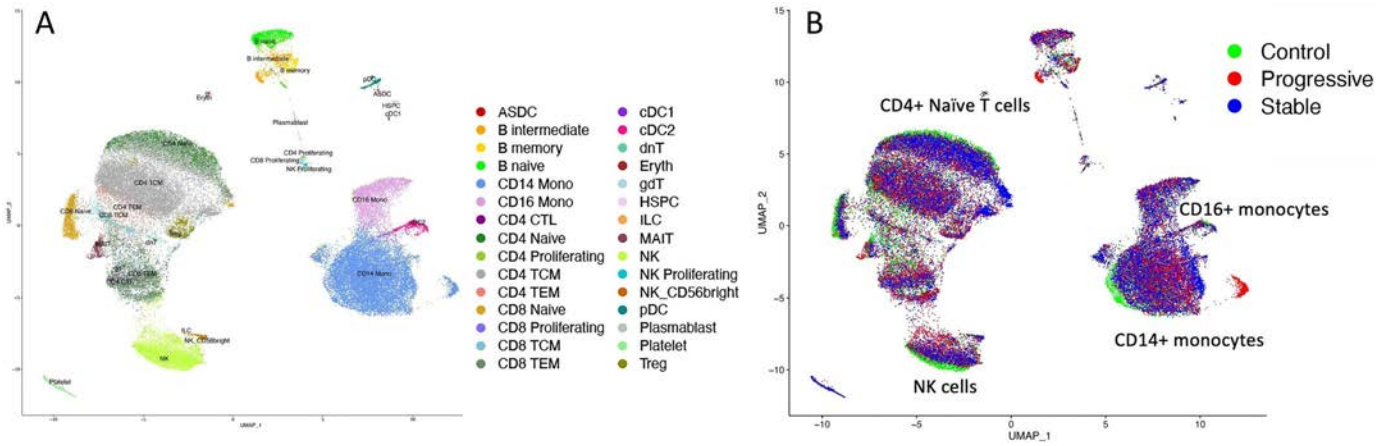

Figure 3. (A) UMAP of PBMC for progressive and stable SSc-ILD patients and controls colored by cell-type. (B) UMAP of peripheral blood NK cells, T cells, and monocytes colored by patient population (eg, progressive disease, stable disease, and control). Shifts within clusters between patients with progressive SSc-ILD and patients with stable SSc-ILD and control participants were observed. ASDC, AXL+Siglec-6+ dendritic cells; cDC, classic dendritic cells; dnT, double negtive T cells; Eryth, erythrocytes; gdT, gamma-delta T cells; HSPC, hemopoietic stem and progenitor cells; ILC, innate lymphoid cells; MAIT, mucosal-associated invariant T; NK, natural killer; PBMC, peripheral blood mononuclear cells; pDC, plasmacytoid dendritic cells; SSc-ILD, scleroderma with interstitial lung disease; TCM, T central memory; TEM, T effector memory cells.

chitinase genes *CHIT1* and *CHI3L1* that encode serum biomarkers of fibrotic lung disease.[13,14] These genes trended toward higher expression in patients with SSc-ILD, *CHI3L1* meeting an uncorrected $P < 0.05$ (Supplementary Figures 6A and 7). Finally, a cluster included genes significantly (*RHOB*, *SAT1*, and *C5AR1*) (Supplementary Figure 6B) or strongly trending higher (*FTH1*, *PLAUR*, and *SERTD3*) in PBMCs from patients with progressive SSc-ILD (Supplementary Figure 6A). Chitinase 1 (CHIT1) and chitinase-3-like protein 1 (CHI3L1) were both much more highly expressed in SPP1 macrophages in patients with SSc-ILD compared with HCs in scRNA-seq data from our previous scRNA-seq datasets from lungs from patients with SSc.[3] A large collection of genes was down-regulated in patients with stable SSc-ILD, including *CRYBB2P1*, the most statistically significantly down-regulated gene (Supplementary Table 4).

Marker genes of progressive SSc-ILD by single-cell RNA-seq of PBMCs. We analyzed PBMCs from HCs (n = 6), patients with stable SSc-ILD (n = 14), or patients with progressive SSc-ILD (n = 13) by scRNA-seq (Figure 3A) and confirmed cluster identities using known marker genes (Supplementary Figure 8). We first looked at the proportions of cells in the different patient populations. Clustering by cell proportion did not reveal any striking associations between the proportions of cells in the different patient populations (Figure 4A). Of the cell populations, CD8+ naive T cells and monocytes (combining CD14+ and CD16+) showed significant differences in cell proportions across the subsets of HCs, patients with stable SSc-ILD, and patients with progressive SSc-ILD (Figure 4B and C). However, these differences were driven by comparisons between the subsets of HCs and patients with SSc-ILD, with significant changes observed between patients with stable SSc-ILD versus HC and patients with progressive SSc-ILD versus HC for CD8+ naive T cells. No significant differences were detected between patients with stable SSc-ILD and patients with progressive SSc-ILD.

Next, we examined gene expression on Uniform Manifold Approximation and Projection (UMAP) plots by sample (not shown) and by disease status (Figure 3B). The latter showed shifts in the cell populations within clusters. Comparing PBMC transcriptomes from HCs with patients with SSc-ILD, two clusters of cells from patients with progressive SSc-ILD stood out separately. However, these clusters were only seen in one patient (not shown) and so were not representative of consistent changes in gene expression associated with patients with progressive SSc-ILD. Closer examination also suggested that there were shifts within clusters between samples from patients with progressive SSc-ILD and patients with stable SSc-ILD, most notably in CD14+ monocytes, CD4+ naive T cells, and NK cells (Figure 4).

DEGs in CD14+ monocytes from patients with progressive SSc-ILD. We systematically examined changes in gene expression associated with cells (scRNA-seq) from patients with progressive SSc-ILD compared with patients with stable SSc-ILD.[15] We then examined data by sample (pseudobulk) to identify biomarkers associated with PBMC cell populations from individual HCs, patients with stable SSc-ILD, and patients with progressive SSc-ILD that might be useful as robust biomarkers of disease progression. CD14+ monocytes showed a series of genes that met criteria for differential expression as well as a difference comparing patients with stable SSc-ILD with patients with progressive SSc-ILD (Supplementary Table 5).

Several also showed statistically significant differences among HC, patients with stable disease, and patients with progressive disease: *LOXHD1*, *ATXN2L*, *EIF4E*, *PIK3CG*, *PRELID1*

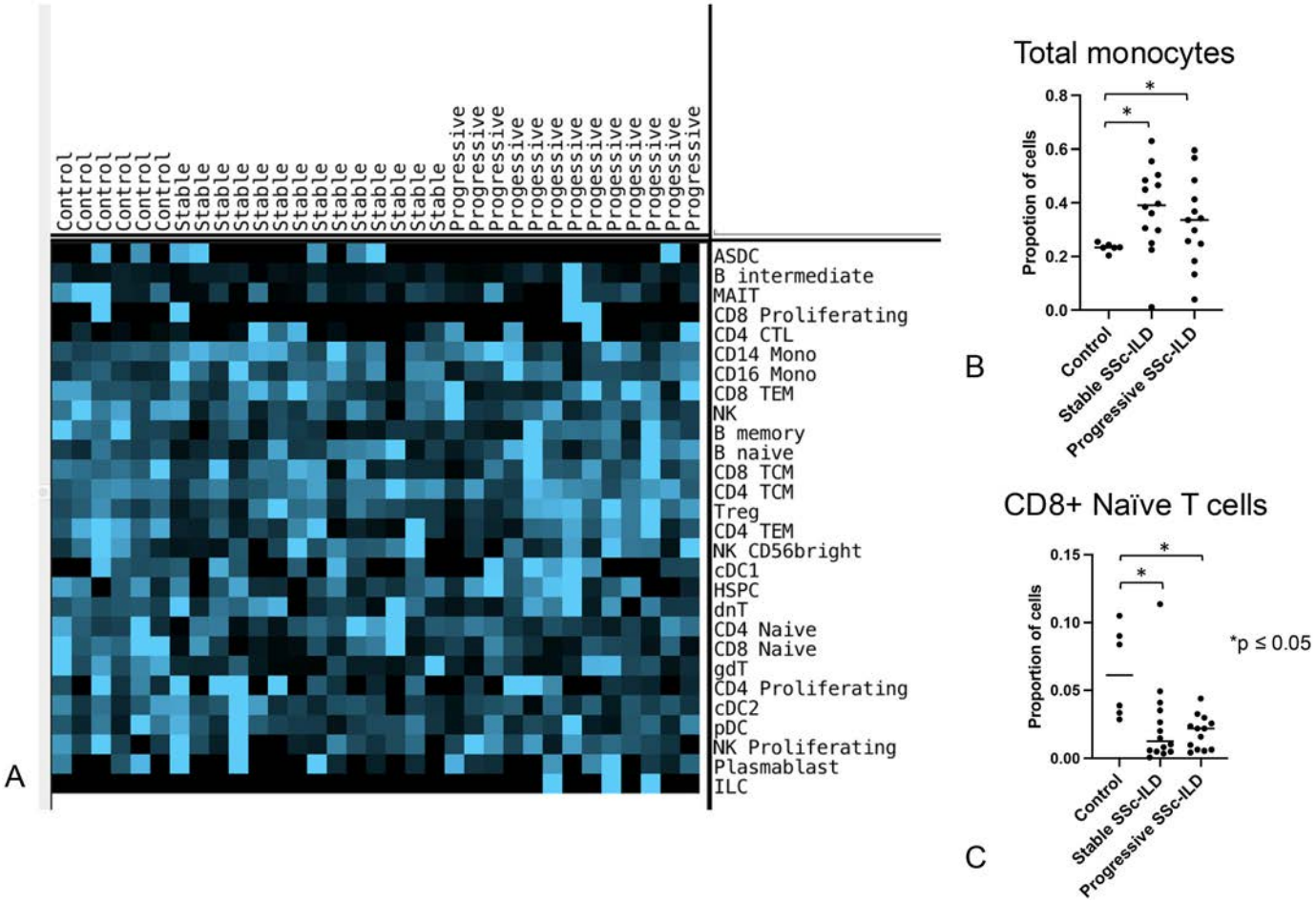

Figure 4. (A) Pseudobulk expression heatmap of cell types according to patient group. (B) Proportion of total monocytes in control, stable SSc-ILD, and progressive SSc-ILD groups each patient population. A significant overall difference was detected ($P = 0.043$) using Kruskal-Wallis test; the asterisk indicates significant difference between control participants and patients with SSc-ILD (stable disease and progressive disease combined) (C) Proportion of CD8+ naive T cells in each patient population. A significant overall difference was detected ($P = 0.015$); post hoc test showed significant differences between control participants and patients with stable SSc-ILD and control participants and patients with progressive SSc-ILD, but not between patients with stable SSc-ILD and patients with progressive SSc-ILD. *$P \leq 0.05$. ASDC, AXL+ Siglec-6+ dendritic cells; cDC, classic dendritic cells; dnT, double negative T cells; gdT, gamma-delta T cells; HSPC, hemopoietic stem and progenitor cells; ILC, innate lymphoid cells; MAIT, mucosal-associated invariant T; NK, natural killer; pDC, plasmacytoid dendritic cells; SSc-ILD, scleroderma with interstitial lung disease; TCM, T central memory cells; TEM, T effector memory cells. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25619/abstract.

(Figure 5A, Supplementary Figure 9) and specific differences between patients with stable disease and patients with progressive disease (Dunn's test). *LOXHD1* was the only gene among these to show differential expression between patients with stable SSc-ILD and patients with progressive SSc-ILD in the bulk RNA-seq data, suggesting a potential role in patients with monocyte-driven progressive SSc-ILD. Although *ATXN2L* did not show significant differences within individual groups, it demonstrated a statistically significant decrease in expression in the bulk RNA-seq dataset between patients with progressive SSc-ILD compared with patients with stable SSc-ILD, suggesting a potential role in disease progression. *ATXN2L* is also down-regulated in all macrophage populations in the lungs of patients with SSc-ILD (Supplementary Figure 10).

Examining genes associated with patients with progressive SSc-ILD in PBMC in bulk RNA-seq, *S100A8*, *S100A9*, and *S100A12*, all highly expressed but not highly up-regulated in the SPP1 macrophages in lungs of patients with SSc-ILD, were not up-regulated in the PBMC CD14+ monocytes of patients with progressive SSc-ILD (not shown). The reason for this difference is uncertain. *CHIT1*, *CHI3L1*, and *MMP9*, although strongly up-regulated in SPP1 macrophages in lungs of patients with SSc-ILD, were not reliably detected in CD14+ monocytes in the scRNA-seq PBMC data (not shown). Thus, *LOXHD1* and *ATXN2L* appeared to be the most consistent biomarkers of progressive SSc-ILD because they were up-regulated (*LOXHD1*) or down-regulated (*ATXN2L*) in CD14+ cells on scRNA-seq and validated in the bulk RNA-seq dataset.

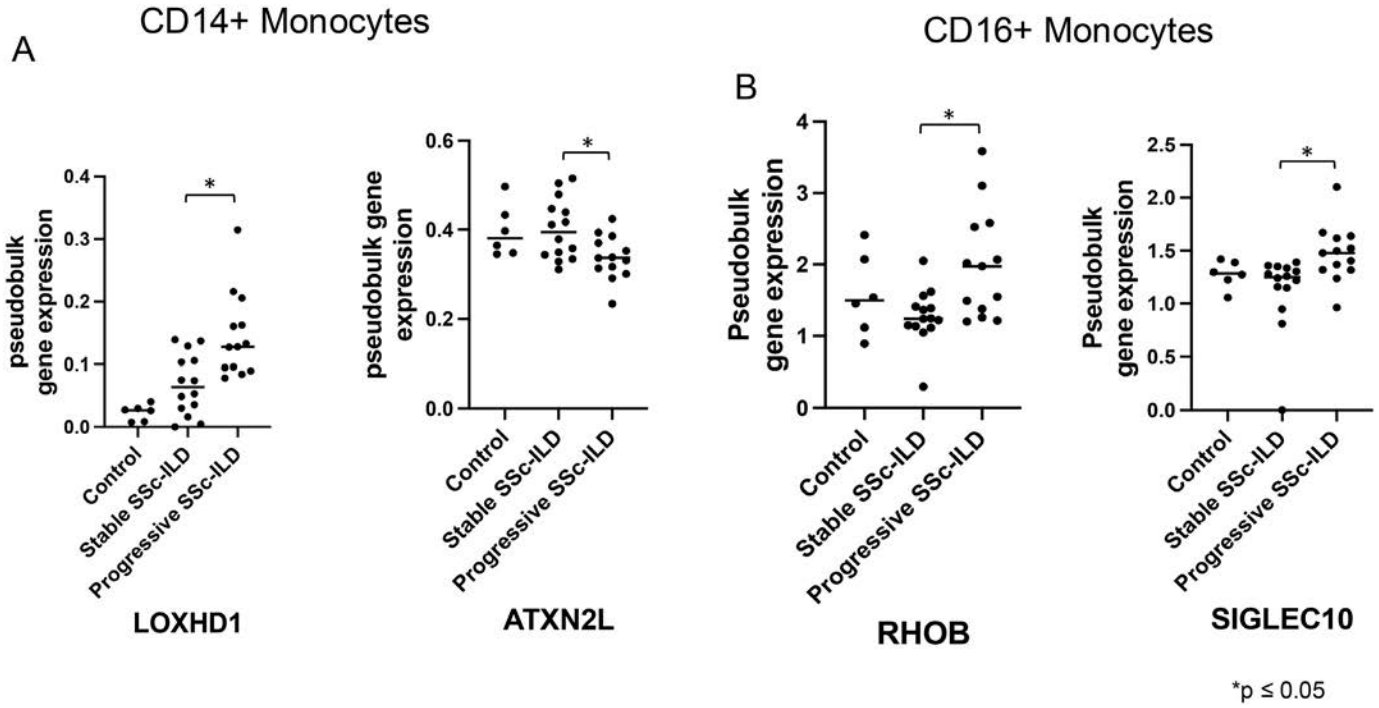

Figure 5. (A) CD14+ monocyte genes (*LOXHD1*, *ATXN2L*) showing significant differences (using MAST and Kruskal-Wallis statistics) in control participants, patients with stable SSc-ILD, and patients with progressive SSc-ILD. (B) CD16+ monocyte genes (*RHOB*, *SIGLEC10*) showing significant differences based on comparing cells (MAST $P_{adj} < 0.05$) and samples ($P < 0.05$, Kruskal-Wallis) in control participants, patients with stable SSc-ILD, and patients with progressive SSc-ILD. *$P \leq 0.05$. SSc-ILD, scleroderma with interstitial lung disease.

We also examined RNA gene changes in patients with SSc-ILD, including both samples from patients with progressive disease and patients with stable disease compared with control PBMCs. We highlight several genes of particular interest. Both *NFKB1* and *RELB*, parts of the classical and alternative NF-κB signaling pathways,[16] were respectively down-regulated in PBMCs from patients with SSc-ILD (Supplementary Figure 11). *NFKB1* is down-regulated in multiple cell populations in the lungs of patients with SSc-ILD (Supplementary Figure 12). *OSM*, *CASS4*, *PRDM1*, and *RAPGEF1* were also down-regulated in lung CD14+ monocytes from patients with SSc-ILD compared with HCs (Supplementary Figure 13).

DEGs in CD16+ monocytes from patients with progressive SSc-ILD. Using a similar approach, we identified a set of genes in CD16+ monocytes associated with patients with progressive SSc-ILD versus patients with stable SSc-ILD in scRNA-seq. Four genes, *RHOB*, *SIGLEC10*, *RHOC*, and *GRK3*, showed consistent differential expression between cells from patients with progressive SSc-ILD and patients with stable SSc-ILD (Figure 5B and Supplementary Figure 14); supported by multiple statistical approaches comparing samples from patients with progressive SSc-ILD with patients with stable SSc-ILD (Figure 5B). Of these genes, only *RHOB* was increased in patients with progressive SSc-ILD in the PBMC bulk RNA-seq data (Figure 5B, Supplementary Figure 15). None of these genes were up-regulated in CD14+ monocytes in the lungs of patients with SSc-ILD compared with control participants. Reexamining the PBMC CD14+ monocyte data, *RHOB* was also increased in CD14+ monocytes from patients with progressive SSc-ILD (*t*-test; $P = 0.028$), although failing to meet statistical significance in the Kruskal-Wallis test. These findings suggest RHOB may serve as a robust peripheral biomarker of disease progression.

Regarding markers identified in CD14+ monocytes, *LOXHD1* showed a trend toward increased expression in CD16+ monocytes but its level of expression was much lower than in CD14+ monocytes. *ATXN2L* failed to show a significant down-regulation in CD16+ monocytes from patents with progressive SSc-ILD. Thus, *RHOB* was the most convincing biomarker of patients with progressive SSc-ILD because it alone was up-regulated in CD16+ and CD14+ cells on scRNA-seq and was validated in the bulk RNA-seq dataset.

DEGs in NK and CD4+ naive T cells from patients with progressive SSc-ILD. NK and CD4+ naive T cells exhibited separation of cells (scRNA-seq) between patients with progressive SSc-ILD and patients with stable SSc-ILD (Figure 3). Eight NK cell genes were significantly differentially expressed between progressive and stable samples (Supplementary Table 5). Among these, three genes—*NCR3*, *RHD*, and *MT-CO2*— also showed significant differences across control participants, patients with progressive SSc-ILD, and patients with stable SSc-ILD,

specifically between samples from patients with progressive SSc-ILD and patients with stable SSc-ILD by Dunn's test (Supplementary Figure 15; Supplementary Table 5). However, none of these were detectably different in the bulk RNA-seq data (Supplementary Figure 16). In prior scRNA-seq of lungs from patients with SSc-ILD, *RHD* was expressed at very low levels but enriched in NK cells; *NCR3* was most highly expressed in lungs from HCs compared with patients with SSc-ILD; and *MT-CO2* was broadly expressed but decreased in lungs from patients with SSc-ILD relative to HCs (Supplementary Figure 16). NK cell effector genes previously reported to be up-regulated in lung NK cells from patients with SSc-ILD—*GZMB*, *IFNG* and *AREG*[6]—were not increased in scRNA-seq from PBMC NK cells or in patients with progressive disease in this dataset (not shown).

Ten naive CD4+ T cell genes were statistically significant when comparing cells from patients with progressive SSc-ILD versus patients with stable SSc-ILD (Supplementary Table 5). Two NK cell genes, *CHI3L2* and *MT-CO2*, also showed significant differences when comparing HCs, patients with progressive SSc-ILD, and patients with stable SSc-ILD; and also between patients with progressive SSc-ILD and patients with stable SSc-ILD by Dunn's test (Supplementary Figure 16). Both genes showed a weak trend toward differences between patients with progressive SSc-ILD and patients with stable SSc-ILD in bulk RNA-seq. As noted above, *MT-CO2* was also significantly decreased in NK cells from PBMCs as well as lungs from multiple populations with SSc-ILD. *CHI3L2* did not show up-regulated expression in lung T cells and was more highly expressed and down-regulated in lung alveolar type 2 cells and myofibroblasts from patients with SSc-ILD (Supplementary Figure 17).

DISCUSSION

Monocyte expression of LOXHD1 stands out as the one gene most highly predictive of progressive SSc-ILD. Most studies of LOXHD1 have focused on its role in hearing loss. It is part of the stereociliary complex protruding from cochlear hair cells and is required for mechanotransduction through membrane structures referred to as tip links. Mutation of LOXHD1 affects mechanotransduction but not the structure of tip links.[17] However, a more recent study suggests another function. Increased expression of an alternative, shorter transcript has been associated with disordered cytoskeleton and decreased metastatic potential.[18]

Our findings are consistent with past studies, indicating that male sex, KL-6, and CRP levels are predictors of progressive disease. KL-6 was seen in one study of a mixed population with SSc-ILD and Mixed Connective Tissue Disease-associated interstitial lung disease (MCTD-ILD) as predictive of progressive disease[19] and in a combined analysis of patients with SSc-ILD.[20] Male sex has previously been identified as a risk factor for progressive SSc-ILD[20] as has CRP levels.[21,22]

In our study cohort, anti-Scl-70 was seen more often in progressive disease, and ACA more commonly in stable SSc-ILD. Anti-Scl-70 has been associated with the presence of SSc-ILD in multiple studies[23–25] and also with progressive SSc-ILD,[10] whereas ACAs were negatively associated with SSc-ILD and trended toward protection from progressive disease.[10,23–25] Another recent study found no relationship between anti-Scl-70 and progressive decline in FVC but did see a decline in patients with anti-Ro antibodies.[26] Variability in studies examining the relationship between anti-Scl-70 and SSc-ILD have been related to the method for its detection.[27] We saw anti-Ro in more patients with progressive SSc-ILD than patients with stable SSc-ILD but was observed too infrequently to be informative about its potential association with progressive disease. Our sample size may have been too small to show statistically significant associations between progressive ILD and these autoantibodies.

Down-regulated expression of ATXN2L and up-regulated expression of RHOB were also markers in patients with progressive SSc-ILD in our scRNA-seq dataset. We validated dysregulated expressions of *ATXN2L* and *RHOB* in the bulk RNA-seq data. Rho proteins are important in regulating cell cytoskeleton through actin fibers, which are also the structural proteins of stereocilia[28] and potentially linking the function of RHOB and LOXHD1. Although RHOB is not required for macrophage chemotaxis,[29] it is tempting to speculate that the role of both LOXHD1 and RHOB might be related to macrophage locomotion. ATXNL2 along with ATXN2 regulate protein translation in circadian rhythm,[30] but ATXNL2 is also part of a macrophage signature that includes SPP1 predicting clinical features in patients with meningioma[31] and is widely expressed by cells in the lungs, suggesting other roles.

S100A12, CHI3L1, and MMP9 are part of clusters or pathways associated with progressive SSc-ILD in bulk RNA-seq but were not up-regulated (S100A12) or not reliably detected (MMP9/CHI3L1) in monocytes in the scRNA-seq dataset. Past serum data strongly support these as PBMC bulk RNA-seq biomarkers. Serum S100A12 is a prognostic biomarker in IPF.[32] Serum levels of CHI3L1 (previously known as YKL-40) are higher in patients with SSc-ILD[33] and patients with IPF.[34] CHI3L1 activates profibrotic macrophages through CRTH2, suggesting a direct role in pathogenesis.[35] MMP9 is elevated in patients with IPF[36] and patients with SSc,[37] but has not been studied specifically in patients with SSc-ILD. Thus, the gene clusters represented by these genes in bulk RNA-seq likely represent true markers of progressive disease.

The origin of SPP1 macrophages in SSc-ILD lungs remains uncertain, ie circulating monocytes or resident lung macrophage populations. Significantly, several genes that show profoundly up-regulated expression by SPP1 macrophages in SSc-ILD lungs (SPP1, CHIT1, CHI3L1 and MMP9);[2,3] are expressed at very low levels by SSc-ILD PBMC monocytes. Studies of patients undergoing bronchoalveolar lavage or retransplant after a transplant

suggest that most alveolar macrophages are donor-derived.[38,39] However, other studies have shown that recipient macrophages largely replace alveolar macrophages after transplant, suggesting that most alveolar macrophages are derived from monocytes.[40] Our previous scRNA-seq data from lungs from patients with SSc-ILD and patients with IPF show that alveolar and profibrotic SPP1-macrophages, but not monocyte-like macrophages are actively proliferating.[2,3] Collectively these observations suggest that monocytes in the periphery do not differentiate in the circulation to profibrotic SPP1 macrophages (and then migrate into the lungs), but that circulating monocytes infiltrate into lungs as monocyte-like macrophages, potentially differentiating into SPP1 macrophages. However, our observations here suggest that either monocytes undergo profound changes in gene expression after migrating into SSc-ILD lung tissues or that the SPP1 macrophages in SSc-ILD derive from proliferation of resident macrophage populations.

CD14+ monocyte counts have been shown as prognostic for transplant-free survival in IPF but not non-IPF fibrotic lung disease (including SSc-ILD).[41,42] Partially consistent with these observations, we detected increased monocyte counts in SSc-ILD compared to control samples, though not showing even a trend toward increased monocytes in patients with progressive SSc-ILD compared with patients with stable SSc-ILD. Although our data suggest that alterations in monocyte counts are associated with the development of SSc-ILD because our control groups did not include a group of SSc patients without ILD, altered monocyte counts could be related to the development of SSc rather than to the development of SSc-ILD. A recent study of monocytes in IPF and hypersensitivity pneumonitis appears to show some common features with altered gene expression in SSc-ILD monocytes, including increased expression of S100A8 and S100A9.[43]

Recently, a group stratified monocyte populations from patients with SSc into three groups based on gene expression.[44] One of these groups (group C) associated with more severe lung disease appears to be similar to the IFN-regulated gene expression we reported previously[7] and appears again in this dataset. A previous study showed that the presence of the IFN signature in patients with SSc PBMC correlates with plasma levels of gene products of two IFN-regulated genes CXCL10 and CXCL11.[45] ITAC/CXCL11 but not IP-10/CXCL10 correlated weakly with FVC. Our dataset here failed to show a significant correlation between progressive SSc-ILD and the IFN signature.

A recent study of proteomic biomarkers of progressive SSc-ILD identified 29 differentially expressed proteins, although not statistically significantly different after adjustment for multiple comparisons. One of these, ZNF500, showed an unadjusted $P < 0.05$ and direction, that is, decreased expression in progressive patients, in our bulk RNA-seq data (unadjusted $P = 0.0148$).

Compared with other longitudinal studies of patients with SSc-ILD, the current study defined patients with progressive disease over a relatively short follow up period. This design identifies patients with progressive disease at the time of blood sampling rather than long-term prognosis. This was felt to be important for making decisions regarding treatment but also felt to be more likely to detect markers of disease activity. In addition, long-term prognosis is increasingly affected by treatments because medications affecting the course of the disease have become available. Notably, the average disease duration was shorter in our patient group with progressive SSc-ILD than in our patient group with stable SSc-ILD despite having lower average baseline FVC. These two observations may be related because previous trajectory analyses have suggested that a subgroup of patients with SSc with low FVC show a rapid decline in FVC.[46] Because our patients were not matched for disease duration or baseline FVC, we cannot exclude the possibility that these variables are affecting the biomarkers we have described. Because this is a relatively small study, it should be viewed as exploratory. For this reason, some comparison between progressive versus regressive SSc-ILD were assessed by uncorrected *P* values. However, these uncorrected *P* values led to significant correlations with pathways, suggesting they detected underlying mechanisms of monocyte activation. Despite study limitations with validation these biomarkers hold promise to assist with the clinically challenging assessment of SSc-ILD.

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 Lafyatis 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

Mitsubishi Inc 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 Mitsubishi Inc.

REFERENCES

1. Lafyatis R, Valenzi E. Assessment of disease outcome measures in systemic sclerosis. Nat Rev Rheumatol 2022;18(9):527–541.
2. Valenzi E, Bulik M, Tabib T, et al. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. Ann Rheum Dis 2019;78(10):1379–1387.
3. Papazoglou A, Huang M, Bulik M, et al. Epigenetic regulation of profibrotic macrophages in systemic sclerosis-associated interstitial lung disease. Arthritis Rheumatol 2022;74(12):2003–2014.
4. Morse C, Tabib T, Sembrat J, et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. Eur Respir J 2019;54(2):1802441.

5. Berman JS, Serlin D, Li X, et al. Altered bleomycin-induced lung fibrosis in osteopontin-deficient mice. Am J Physiol Lung Cell Mol Physiol 2004;286(6):L1311–L1318.
6. Padilla CM, Valenzi E, Tabib T, et al. Increased CD8+ tissue resident memory T cells, regulatory T cells, and activated natural killer cells in systemic sclerosis lungs. Rheumatology (Oxford) 2024;63(3):837–845.
7. York MR, Nagai T, Mangini AJ, et al. A macrophage marker, Siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by Type I interferons and toll-like receptor agonists. Arthritis Rheum 2007;56(3):1010–1020.
8. Christmann RB, Hayes E, Pendergrass S, et al. Interferon and alternative activation of monocyte/macrophages in systemic sclerosis-associated pulmonary arterial hypertension. Arthritis Rheum 2011; 63(6):1718–1728.
9. McNearney TA, Reveille JD, Fischbach M, et al. Pulmonary involvement in systemic sclerosis: associations with genetic, serologic, sociodemographic, and behavioral factors. Arthritis Rheum. 2007; 57(2):318–326.
10. Assassi S, Sharif R, Lasky RE, et al; GENISOS Study Group. Predictors of interstitial lung disease in early systemic sclerosis: a prospective longitudinal study of the GENISOS cohort. Arthritis Res Ther 2010;12(5):R166.
11. Richards TJ, Kaminski N, Baribaud F, et al. Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2012;185(1):67–76.
12. Craig VJ, Zhang L, Hagood JS, et al. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. Am J Respir Cell Mol Biol 2015;53(5):585–600.
13. Lee CG, Herzog EL, Ahangari F, et al. Chitinase 1 is a biomarker for and therapeutic target in scleroderma-associated interstitial lung disease that augments TGF-β1 signaling. J Immunol 2012;189(5): 2635–2644.
14. Majewski S, Szewczyk K, Jerczyńska H, et al. Longitudinal and comparative measures of serum chitotriosidase and YKL-40 in patients with idiopathic pulmonary fibrosis. Front Immunol 2022;13:760776.
15. Finak G, McDavid A, Yajima M, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol 2015; 16:278.
16. Dimitrakopoulos FID, Antonacopoulou AG, Kottorou AE, et al. Expression of intracellular components of the NF-κB alternative pathway (NF-κB2, RelB, NIK and Bcl3) is associated with clinical outcome of NSCLC patients. Sci Rep 2019;9(1):14299.
17. Trouillet A, Miller KK, George SS, et al. *Loxhd1* mutations cause mechanotransduction defects in cochlear hair cells. J Neurosci 2021;41(15):3331–3343.
18. Deng Q, Natesan R, Cidre-Aranaz F, et al. Oncofusion-driven de novo enhancer assembly promotes malignancy in Ewing sarcoma via aberrant expression of the stereociliary protein LOXHD1. Cell Rep 2022; 39(11):110971.
19. Yamakawa H, Hagiwara E, Kitamura H, et al. Serum KL-6 and surfactant protein-D as monitoring and predictive markers of interstitial lung disease in patients with systemic sclerosis and mixed connective tissue disease. J Thorac Dis 2017;9(2):362–371.
20. Winstone TA, Assayag D, Wilcox PG, et al. Predictors of mortality and progression in scleroderma-associated interstitial lung disease: a systematic review. Chest 2014;146(2):422–436.
21. Yoshida Y, Sugimoto T, Hosokawa Y, et al. C-reactive protein and ground-glass opacity as predictors for intractable interstitial lung disease in patients with systemic sclerosis under cyclophosphamide treatment regardless of concomitant glucocorticoids. Mod Rheumatol 2022;32(1):141–148.
22. Liu X, Mayes MD, Pedroza C, et al. Does C-reactive protein predict the long-term progression of interstitial lung disease and survival in patients with early systemic sclerosis? Arthritis Care Res (Hoboken). 2013;65(8):1375–1380.
23. Steen VD. Autoantibodies in systemic sclerosis. Semin Arthritis Rheum 2005;35(1):35–42.
24. Iniesta Arandia N, Simeón-Aznar CP, Guillén Del Castillo A, et al; RESCLE investigators, Autoimmune Diseases Study Group (GEAS). Influence of antibody profile in clinical features and prognosis in a cohort of Spanish patients with systemic sclerosis. Clin Exp Rheumatol 2017;35 Suppl 106(4):98–105.
25. Liaskos C, Marou E, Simopoulou T, et al. Disease-related autoantibody profile in patients with systemic sclerosis. Autoimmunity 2017; 50(7):414–421.
26. Hamberg V, Sohrabian A, Volkmann ER, et al. Anti-Ro52 positivity is associated with progressive interstitial lung disease in systemic sclerosis-an exploratory study. Arthritis Res Ther 2023;25(1):162.
27. Jandali B, Salazar GA, Hudson M, et al. The effect of anti-Scl-70 antibody determination method on its predictive significance for interstitial lung disease progression in systemic sclerosis. ACR Open Rheumatol 2022;4(4):345–351.
28. Flock A, Cheung HC. Actin filaments in sensory hairs of inner ear receptor cells. J Cell Biol 1977;75(2 Pt 1):339–343.
29. Königs V, Jennings R, Vogl T, et al. Mouse macrophages completely lacking Rho subfamily GTPases (RhoA, RhoB, and RhoC) have severe lamellipodial retraction defects, but robust chemotactic navigation and altered motility. J Biol Chem 2014;289(44):30772–30784.
30. Zhuang Y, Li Z, Xiong S, et al. Circadian clocks are modulated by compartmentalized oscillating translation. Cell 2023;186(15):3245–3260.e23.
31. Zhang X. Single-cell RNA sequencing identifies macrophage signatures correlated with clinical features and tumour microenvironment in meningiomas. IET Syst Biol 2023;17(5):259–270.
32. Lu Y, Chen J, Tang K, et al. Development and validation of the prognostic index based on inflammation-related gene analysis in idiopathic pulmonary fibrosis. Front Mol Biosci 2021;8:667459.
33. Nordenbaek C, Johansen JS, Halberg P, et al. High serum levels of YKL-40 in patients with systemic sclerosis are associated with pulmonary involvement. Scand J Rheumatol 2005;34(4):293–297.
34. Furuhashi K, Suda T, Nakamura Y, et al. Increased expression of YKL-40, a chitinase-like protein, in serum and lung of patients with idiopathic pulmonary fibrosis. Respir Med 2010;104(8):1204–1210.
35. Cao Y, Rudrakshala J, Williams R, et al. CRTH2 mediates profibrotic macrophage differentiation and promotes lung fibrosis. Am J Respir Cell Mol Biol 2022;67(2):201–214.
36. Todd JL, Vinisko R, Liu Y, et al; IPF-PRO Registry Investigators. Circulating matrix metalloproteinases and tissue metalloproteinase inhibitors in patients with idiopathic pulmonary fibrosis in the multicenter IPF-PRO Registry cohort. BMC Pulm Med 2020;20(1):64.
37. Kim WU, Min SY, Cho ML, et al. Elevated matrix metalloproteinase-9 in patients with systemic sclerosis. Arthritis Res Ther 2005;7(1): R71–R79.
38. Nayak DK, Zhou F, Xu M, et al. Long-term persistence of donor alveolar macrophages in human lung transplant recipients that influences donor-specific immune responses. Am J Transplant 2016;16(8): 2300–2311.
39. Eguíluz-Gracia I, Schultz HH, Sikkeland LI, et al. Long-term persistence of human donor alveolar macrophages in lung transplant recipients. Thorax. 2016;71(11):1006–1011.
40. Byrne AJ, Powell JE, O'Sullivan BJ, et al. Dynamics of human monocytes and airway macrophages during healthy aging and after transplant. J Exp Med 2020;217(3):e20191236.

41. Scott MKD, Quinn K, Li Q, Carroll R, et al. Increased monocyte count as a cellular biomarker for poor outcomes in fibrotic diseases: a retrospective, multicentre cohort study. Lancet Respir Med 2019;7(6): 497–508.
42. Kreuter M, Lee JS, Tzouvelekis A, et al. Monocyte count as a prognostic biomarker in patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2021;204(1):74–81.
43. Zhao AY, Unterman A, Abu Hussein N, et al. Peripheral blood single-cell sequencing uncovers common and specific immune aberrations in fibrotic lung diseases. bioRxiv 2023. https://doi.org/10.1101/2023.09.20.558301
44. Makinde HM, Dunn JLM, Gadhvi G, et al. Three distinct transcriptional profiles of monocytes associate with disease activity in scleroderma patients. Arthritis Rheumatol 2023;75(4):595–608.
45. Liu X, Mayes MD, Tan FK, et al. Correlation of interferon-inducible chemokine plasma levels with disease severity in systemic sclerosis. Arthritis Rheum. 2013;65(1):226–235.
46. Man A, Davidyock T, Ferguson LT, Ieong M, Zhang Y, Simms RW. Changes in forced vital capacity over time in systemic sclerosis: application of group-based trajectory modelling. Rheumatology (Oxford) 2015;54(8):1464–1471.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 469–477
DOI 10.1002/acr.25647

American College of Rheumatology
Empowering Rheumatology Professionals

The Effect of Pulmonary Function Test Reference Equations on Classification of Restrictive Lung Disease Severity in Systemic Sclerosis

Kamini E. Kuchinad, Rachel S. Wallwork, Matthew R. Lammi, Christopher A. Mecoli, Julie J. Paik, Fredrick M. Wigley, Robert A. Wise, Laura K. Hummers, Ami A. Shah, and Ji Soo Kim

Objective. Interstitial lung disease (ILD) is a significant cause of morbidity and mortality in systemic sclerosis (SSc), particularly among Black patients. Pulmonary function tests (PFTs) are critical to screen for and monitor SSc-ILD. We examined whether race-specific and race-neutral PFT reference equations impact classification of restrictive lung disease (RLD) severity in Black and White patients with SSc.

Methods. Baseline percent predicted forced vital capacity (ppFVC) was calculated for self-identified Black (n = 641) and White (n = 2,909) patients in the Johns Hopkins Scleroderma Center Research Registry using race-specific (Global Lung Function Initiative 2012 [GLI 2012] and National Health and Nutrition Examination Survey III) and race-neutral (GLI Global) equations. The percentage of Black and White individuals who switched RLD severity categories (normal [ppFVC ≥ 80%], mild [70% ≤ ppFVC < 80%], moderate [60% ≤ ppFVC < 70%], severe [50% ≤ ppFVC < 60%], or very severe [ppFVC < 50%]) with the use of race-neutral versus race-specific equations was calculated. The percentages of patients who would meet typical ppFVC thresholds for immunosuppression, clinical trial eligibility, and lung transplant referral were compared.

Results. Black individuals had lower absolute FVC values than White individuals. Forty-seven percent (n = 303) of Black individuals were reclassified as having more severe RLD, and 17% (n = 487) of White individuals were reclassified as having less severe RLD when using the GLI Global versus GLI 2012 equations. Statistically greater proportions of Black individuals met ppFVC thresholds for immunosuppression, clinical trial eligibility, and lung transplant referral with race-neutral versus race-specific equations.

Conclusion. The use of race-specific PFT reference equations may result in systematic misclassification of ILD severity with potential impact on health care delivery and clinical trial eligibility.

INTRODUCTION

Systemic sclerosis (SSc) is a complex autoimmune disease characterized by fibrosis, vasculopathy, and immune dysregulation. Individuals who self-identify as Black have higher incidence of SSc and develop SSc at a younger age than their White counterparts.[1] Furthermore, Black individuals with SSc face higher rates of interstitial lung disease (ILD), severe restrictive lung disease (RLD), and overall mortality as compared to White individuals with SSc.[1–3] Though the etiology of these differences is not well understood, previous studies have suggested that it is multifactorial, with differences in genetic background, immune responses, and socioeconomic status likely playing a role.[4–10]

Examining the tools we use to diagnose and monitor SSc-ILD is important in elucidating these differences in health outcomes. Pulmonary function tests (PFTs), particularly assessment

Presented in part at the American College of Rheumatology Convergence 2024, Washington, DC, November 17, 2024.

Supported in part by the Jerome L. Greene Foundation, the Johns Hopkins inHealth Initiative, 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, K24-AR-080217, and T32-AR-048522-19).

Kamini E. Kuchinad, MD, MPH, Rachel S. Wallwork, MD, MHS, Matthew R. Lammi, MD, MSCR, Christopher A. Mecoli, MD, MHS, Julie J. Paik, MD, MHS, Fredrick M. Wigley, MD, Robert A. Wise, MD, Laura K. Hummers, MD, ScM, Ami A. Shah, MD, MHS, Ji Soo Kim, PhD: Johns Hopkins University School of Medicine, Baltimore, Maryland.

Drs Shah and Kim contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25647).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25647.

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

Submitted for publication February 26, 2025; accepted in revised form August 19, 2025.

SIGNIFICANCE & INNOVATIONS

- This study demonstrates that the use of race-specific (Global Lung Function Initiative 2012 [GLI 2012]; National Health and Nutrition Examination Survey III) versus race-neutral (GLI Global) pulmonary function test (PFT) reference equations leads to clinically significant differences in classification of restrictive lung disease severity among Black and White individuals with systemic sclerosis (SSc).
- This study shows that more Black individuals with SSc would meet clinical percent predicted forced vital capacity (ppFVC) thresholds for immunosuppression initiation, lung transplant referral, and clinical trial eligibility with the use of race-neutral versus race-specific equations.
- Rheumatologists should be aware of the ways in which PFT reference equations may evolve over time, as transitioning between different equations may result in shifts in ppFVC that are not reflective of true clinical change.

of the percent predicted forced vital capacity (ppFVC), are used to screen for and monitor SSc-ILD. The American College of Rheumatology (ACR) guidelines for screening for and monitoring ILD in people with systemic autoimmune rheumatic disease recommend that individuals with SSc should undergo regular screening with PFTs and that individuals with diagnosed SSc-ILD should undergo monitoring with PFTs every three to six months at the beginning of their illness.[11] Thus, in clinical practice, individuals with SSc undergo regular monitoring with PFTs at least yearly but often more frequently. The ppFVC is a key measure in identifying RLD; a ppFVC of less than 80% or a decline in ppFVC of greater than 10% is highly concerning for RLD and thus ILD among individuals with SSc. As such, the accuracy of ppFVC is of critical importance in clinical decision-making in the context of diagnosis and management of ILD. Decisions about whether to initiate immunosuppressive or antifibrotic therapy, clinical trial eligibility, and lung transplant referral factor in the ppFVC as a surrogate for ILD severity.

Historically, predicted values for FVC were calculated using equations in the National Health and Nutrition Examination Survey III (NHANES III) or the Global Lung Function Initiative 2012 (GLI 2012), which included age, sex, race, and height as inputs.[12–16] However, the use of race as a factor in PFT reference equations lacks a biologic basis and thus has been questioned.[17,18] Previous studies suggest that race-specific equations may not be accurate predictors of lung pathology and may result in underdiagnosis and undertreatment of disease in Black patients.[17,19–24] In this setting, a race-neutral set of equations known as the GLI Global equations have been developed.[25,26] The American Thoracic Society now recommends the use of these reference equations in characterization of PFTs.[18]

The impact of race-specific and race-neutral equations on RLD classification in Black and White patients with SSc has not been evaluated. Importantly, evidence evaluating risk factors for and management of SSc-ILD has been based on data using race-specific equations for calculation of ppFVC, highlighting the need for re-evaluation in the context of accurately defined outcomes. This study aims to examine the impact of race-specific (NHANES III, GLI 2012) and race-neutral (GLI Global) reference equations on the classification of RLD and health care delivery for self-identified Black and White individuals with SSc via analysis of a large single-center SSc cohort.

PATIENTS AND METHODS

Study population. Established in 1990, the Johns Hopkins Scleroderma Center Research Registry (JHSC) contains sociodemographic, clinical, and serologic characteristics from a population of >4,400 individuals with SSc. This study analyzed routinely collected data and had local ethics committee approval. Adult patients (≥18 years of age) who had at least one PFT, including spirometry measurements, and self-identified as Black or White were included. All individuals met the 1980 or 2013 ACR classification criteria for SSc; had at least three of five features of CREST syndrome (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia); or had definite Raynaud phenomenon, abnormal nailfold capillaries, and a scleroderma-specific antibody.[27,28] PFTs were captured from testing at various sites, including our academic medical center and local testing sites. We performed sensitivity analyses by using a restricted sample of patients who met the 2013 ACR classification criteria.

Scleroderma characteristics. Clinical and demographic characteristics were obtained from the registry. Participants self-reported race and sex, choosing from a fixed set of categories. Age at disease onset was defined as the age at the first non–Raynaud phenomenon SSc symptom. Disease duration refers to the time between the first non–Raynaud's SSc symptom and the first available PFT. History of smoking was a binary variable, with individuals characterized as "ever" versus "never" smokers. Autoantibody testing was performed in patients who consented to provide a serum sample for research using the EuroBlotOne platform (Euroline Systemic Sclerosis Profile, Euroimmun Diagnostics). This assay tests for antibodies against centromere, RNA polymerase III (referred to as POLR3), Scl-70, PM/Scl, NOR-90, Th/To, Ku, U3 RNP, Ro52, and PDGFR. Anti-centromere, anti-POLR3, and anti-PM/Scl testing included testing for two subunits each, and patients were considered positive if both subunits were positive. Serum samples were assayed for anti–U1 RNP antibodies using a commercial ELISA (Inova Diagnostics). ILD was defined as the presence of fibrosis on chest radiography or computed tomography imaging of the chest or the

presence of crackles on examination. Pulmonary hypertension (PH) was defined as a right ventricular systolic pressure of greater than 45 mm Hg on the echocardiogram.[29] Criteria for myopathy included abnormal muscle enzymes, myopathic EMG, abnormal muscle MRI, or abnormal muscle biopsy. Clinical characteristics were considered present if criteria were met at any time during follow-up. To test whether the distributions of variables are similar in Black and White patients, we used two-sample *t*-tests assuming unequal variances for continuous variables and chi-square tests for categorical variables. *P* values are adjusted using the Benjamini-Hochberg method to control for the false discovery rate.[30]

PFT reference equations. Three sets of reference equations were examined. In this section, we use terminology for racial groups as defined in the original articles for the development of the NHANES III and GLI 2012. The NHANES III equations, developed in 1999, calculate predicted FVC values using separate reference equations for Caucasian, African American, and Mexican American adults and youths for each sex.[12] The predicted normal FVC values are calculated as a function of age and height. The GLI 2012 equations also calculate predicted FVC values based on age, height, sex, and race.[13] Reference equations were derived for ethnic and racial categories defined as Caucasian, African American, and North East and South East Asian. In the GLI 2012, log-transformed FVC is modeled as a function of log-transformed height and age, a smoothing age spline, and racial group. Individuals who do not identify themselves with any of these four racial groups or are of mixed ethnic origin are categorized into an "other" group, which serves as a reference for calculations of normal FVC. The GLI Global equation is a single race-neutral equation developed by using an inverse probability weight assigned to each observation of the GLI spirometry data based on race and stratified by sex.[25] The general form of the equation is identical to that of the GLI 2012, but race is no longer used as an input for calculations. For each of the three approaches, a ppFVC value is obtained by dividing the measured FVC value by the predicted normal FVC value and is expressed as a percentage.

ppFVC values were calculated for Black and White individuals from the earliest PFT available in the JHSC using race-specific (NHANES III, GLI 2012) and race-neutral (GLI Global) equations. Age and height captured at the time of the first-available PFT were used to calculate the predicted FVC. Self-identified race was used for the NHANES III and GLI 2012 equations.

Based on the American Thoracic Society recommendations for interpretation of spirometry, we additionally converted raw FVC values to z scores based on NHANES III, GLI 2012, and GLI Global equations.[15] Raw FVC values, age, height, and sex were used to calculate z scores using GLI Global equations. z scores using NHANES III and GLI 2012 equations additionally required ethnicity or race.

Percent predicted diffusing capacity for carbon monoxide (ppDL$_{CO}$) was calculated using GLI 2017 equations, as these are the latest equations available.[31] Notably, these reference equations were only available for White individuals.

Outcome measures. To study the general impact of using race-specific and race-neutral equations on different age, sex, and race groups, we plotted ppFVC by age and sex using the NHANES III, GLI 2012, and GLI Global equations. The average raw FVC, height, and age in 10-year age brackets were used to calculate the groupwise ppFVC. This approach allows us to visualize the difference in the ppFVC that is entirely attributable to using race-specific equations compared to using race-neutral equations for an average person with the same FVC, height, sex, and age in each age bracket.

For each patient, baseline lung function was classified as being normal (≥80% ppFVC) or as being indicative of mild (70%–79% ppFVC), moderate (60%–69% ppFVC), severe (50%–59% ppFVC), or very severe (<50% ppFVC) RLD. The number and proportion of individuals who switched classification categories with the use of the race-neutral (GLI Global) and race-specific (GLI 2012 or NHANES III) equations were identified.

Using z scores, lung function was categorized into normal (z score > −1.645) or mild (−2.5 ≤ z score ≤ −1.645), moderate (−4 ≤ z score < −2.5), or severe (z score < −4) RLD.[15] We used Wilcoxon signed rank tests to test whether there is a nonzero difference between lung function classification when using the race-neutral equations compared to using race-specific equations for all patients and separately for each race. Contingency tables of RLD categories defined by race-neutral and race-specific equations were generated. We used clinically relevant ppFVC thresholds for immunosuppressive medication initiation, clinical trial eligibility, and lung transplant referral as a theoretical exercise to assess the potential impact of different PFT reference equations on medical decision-making. The ppFVC thresholds for clinical trial eligibility were based on the inclusion criteria for the Scleroderma Lung Study I (SLS I) and SLS II; these studies included individuals with a ppFVC between 45% and 79%.[32,33] In clinical practice, the decision to initiate immunosuppressive therapy is often determined by varying thresholds of RLD or a significant decline in ppFVC (≥10%); in our study, we used a conservative threshold of having at least moderate RLD, defined as ppFVC < 70%.[34,35] We used a conservative threshold of ppFVC ≤ 50% to identify patients who would be candidates for lung transplant referral. This threshold was based on the average ppFVC of patients with SSc-ILD who underwent lung transplants in a European multicenter cohort study.[36] For each race, McNemar tests were used to compare the proportion of individuals who

met each clinical criteria using race-neutral and race-specific equations.

RESULTS

This study included 2,909 self-identified White and 641 self-identified Black individuals from the JHSC. Demographic characteristics of the two groups are shown in Table 1. Black individuals had an earlier age at SSc onset (41.0 vs 43.7 years; $P < 0.0001$), shorter disease duration (5.4 vs 6.6 years; $P = 0.0005$), and younger age at the time of spirometry (48.3 vs 54.4 years; $P < 0.0001$) compared to White individuals. Diffuse cutaneous SSc was more common among Black individuals, whereas limited cutaneous SSc was most common among White individuals. The autoantibody distributions of the two groups were distinct; anticentromere, anti–RNA polymerase III, and anti-PM/Scl were enriched among White individuals, and anti–U1 RNP, antifibrillarin, anti-Scl-70, anti-Ku, and anti-Ro52 were more common among Black individuals. Both groups were predominantly female. Black individuals had more severe lung disease on initial PFTs compared to White individuals, with a lower absolute value of FVC (2.4 vs 3.0 L; $P < 0.0001$) and lower total lung capacity (3.8 vs 4.8 L; $P < 0.0001$). Black individuals had lower baseline DLCO (12.9 vs 16.0 mL/min/mm Hg; $P < 0.0001$) and ppDLCO as calculated by the GLI 2017 equations (57.4% vs 73.9%; $P < 0.0001$). In addition, echocardiographic evidence of PH (48.5% vs 33.5%; $P < 0.0001$) and myopathy (30.7% vs 14.5%; $P < 0.0001$) was more common during the disease course in Black

Table 1. Demographics of Black and White individuals with systemic sclerosis from the Johns Hopkins Scleroderma Center Research Registry*

Variable	White (n = 2,909)	Black (n = 641)	*P* value
Age (at time of spirometry), mean (SD), y	54.4 (13.3)	48.3 (13.8)	<0.0001
Age at onset, mean (SD), y	43.7 (15.1); n = 2,772	41.0 (14.2); n = 598	<0.0001
Sex, n (%)			0.8759
Female	2,404 (82.6)	527 (82.2)	
Male	505 (17.4)	114 (17.8)	
Smoking (ever), n (%)	1,385 (48.5); n = 2,856	240 (38.3); n = 626	<0.0001
Disease duration, mean (SD), y	6.6 (8.3); n = 2,852	5.4 (7.2); n = 633	0.0005
Cutaneous type, n (%)			<0.0001
Sine	204 (7.0); n = 2,906	37 (5.8)	
Limited	1,723 (59.3); n = 2,906	250 (39.0)	
Diffuse	979 (33.7); n = 2,906	354 (55.2)	
Height, mean (SD), m	1.65 (0.09)	1.66 (0.09)	0.0195
FEV_1, mean (SD), L	2.4 (0.7); n = 2,903	1.9 (0.6); n = 640	<0.0001
FVC, mean (SD), L	3.0 (0.9)	2.4 (0.7)	<0.0001
DLCO, mean (SD), mL/min/mm Hg	16.0 (6.0); n = 2,668	12.9 (5.8); n = 588	<0.0001
ppDLCO, mean (SD), %[a]	73.9 (24.2); n = 2,668	57.4 (23.9); n = 588	<0.0001
TLC, mean (SD), L	4.8 (5.4); n = 2,509	3.8 (1.2); n = 528	<0.0001
ILD, n (%)	1,082 (40.2); n = 2,691	362 (60.8); n = 595	<0.0001
PH, n (%)	737 (33.5); n = 2,198	234 (48.5); n = 482	<0.0001
Myopathy, n (%)	409 (14.5); n = 2,818	189 (30.7); n = 616	<0.0001
Scleroderma renal crisis, n (%)	120 (4.1)	34 (5.3)	0.2517
Antibody, n (%)			
Centromere	867 (32.0); n = 2,709	53 (8.9); n = 595	<0.0001
Scl-70	453 (19.2); n = 2,354	154 (29.4); n = 523	<0.0001
POLR3	428 (19.1); n = 2,240	59 (12.1); n = 489	0.0004
Fibrillarin	84 (4.0); n = 2,118	91 (20.2); n = 450	<0.0001
Ku	69 (3.3); n = 2,118	47 (10.4); n = 450	<0.0001
NOR-90	85 (4.0); n = 2,118	20 (4.4); n = 450	0.8374
PM/Scl	64 (3.0); n = 2,118	2 (0.4); n = 450	0.0038
Ro52	564 (26.6); n = 2,118	144 (32.0); n = 450	0.0283
Th/To	184 (8.7); n = 2,118	39 (8.7); n = 450	1.0000
U1 RNP	171 (9.2); n = 1,853	111 (27.1); n = 410	<0.0001

* Disease onset was defined as first non–Raynaud phenomenon symptom. Disease duration was defined as years since onset to first PFT. ILD was defined as the presence of fibrosis on chest radiography or computed tomography imaging of the chest or the presence of crackles on examination. PH was defined as a right ventricular systolic pressure of greater than 45 mm Hg on the echocardiogram. Criteria for myopathy included abnormal muscle enzymes, a myopathic electromyogram, abnormal muscle magnetic resonance imaging results, or abnormal muscle biopsy findings. Two-sample *t*-tests assuming unequal variances were used for continuous variables, and chi-square tests were used for categorical variables. The presented *P* values were adjusted using the Benjamini-Hochberg method to control for the false discovery rate.[30] DLCO, diffusing capacity for carbon monoxide; FEV_1, forced expiratory vital capacity in 1 second; FVC, forced vital capacity; ILD, interstitial lung disease; PFT, pulmonary function test; PH, pulmonary hypertension; POLR3, RNA polymerase III antibody; ppDLCO, percent predicted DLCO; TLC, total lung capacity.

[a] Calculated using the Global Lung Function Initiative 2017 formula at baseline.

patients than in White patients, whereas a history of smoking was less common (38.3% vs 48.5%; $P < 0.0001$).

Figure 1 illustrates the differences in calculation of ppFVC by age decile using the NHANES III, GLI 2012, and GLI Global equations. When comparing two individuals of the same sex, height, and absolute value FVC, the race-specific Black equations (NHANES III, GLI 2012) result in a higher ppFVC value compared to the race-specific White equations. Among Black patients, the use of race-specific equations results in a higher calculated ppFVC across age groups compared to the race-neutral (GLI Global) equations for both male and female individuals; in contrast, race-specific equations result in lower calculated ppFVC across age groups and sex strata for White patients compared to race-neutral equations.

Figure 2 demonstrates how the use of race-specific versus race-neutral reference equations affects RLD classification for Black and White individuals. The shaded regions demonstrate the percentage of individuals who switched severity categories with the use of different equations. In total, 47% (n = 303) of Black individuals were reclassified to more severe RLD categories when using the GLI Global versus the GLI 2012 equations. Fifteen percent of Black individuals with normal lung function based on ppFVC (ppFVC ≥ 80%) as per GLI 2012 equations had mildly restricted lung volumes (ppFVC of 70%–79%) when calculated with the GLI Global equations. In contrast, the use of the GLI Global equations resulted in a shift toward less severe RLD for White individuals as compared to the GLI 2012 equations. Seventeen percent (n = 487) of White individuals were reclassified to less severe RLD categories with the use of the GLI Global equations compared to the GLI 2012 equations. Eight percent of White individuals who were classified as having mild RLD (ppFVC of 70%–79%) as per GLI 2012 equations had normal calculated ppFVC values when using the GLI Global equations. Similar trends were observed when comparing the NHANES III and GLI Global equations (Supplemental Figure 1); 53% of Black individuals were classified as having more severe RLD and 25% of White individuals were classified as having less severe RLD when using the GLI Global versus NHANES III equations.

Table 2 demonstrates how the use of race-specific versus race-neutral equations affects the number of individuals who met our previously defined thresholds for clinical trial eligibility, initiation of immunosuppression, and lung transplant referral. When comparing the GLI Global equations to the NHANES III and GLI 2012 equations, an increased proportion of Black individuals met commonly used ppFVC thresholds for clinical trial eligibility (69% vs 54% [$P < 0.0001$] and 69% vs 57%, [$P < 0.0001$]), initiation of immunosuppression (56% vs 39% [$P < 0.0001$] and 56% vs 40% [$P < 0.0001$]), and lung transplant referral (14% vs 8% [$P < 0.0001$] and 14% vs 9% [$P < 0.0001$]). Notably, fewer White individuals met these clinical thresholds with the use of the GLI Global equations as compared to the NHANES III or GLI 2012 equations. There were no qualitative changes in the findings when we restricted the analysis to patients who met the 2013 ACR classification criteria (data not shown).

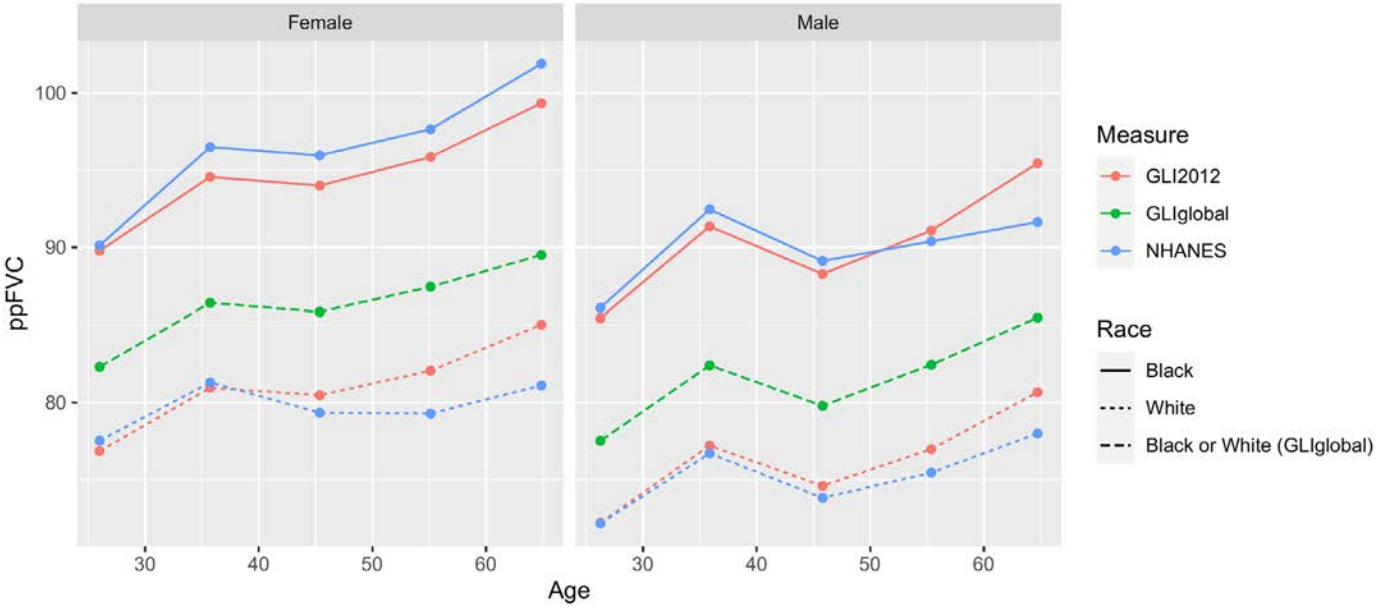

Figure 1. The effect of race-specific versus race-neutral equations on the calculation of ppFVC. The average ppFVC was calculated in 10-year age brackets using race-specific (NHANES III, GLI 2012) and race-neutral (GLI Global) equations for male and female individuals with SSc. For each age bracket, ppFVC was calculated using the same average FVC and height based on either a race-neutral or race-specific equation. The solid lines represent calculation of ppFVC with the GLI 2012 and NHANES III equations for Black race. The lines with small dashes represents ppFVC calculated using the GLI 2012 and NHANES III equations for White race. The green line with larger dashes represents ppFVC as calculated by the GLI Global equations. GLI, Global Lung Function Initiative; NHANES III, National Health and Nutrition Examination Survey III; ppFVC, percent predicted forced vital capacity; SSc, systemic sclerosis.

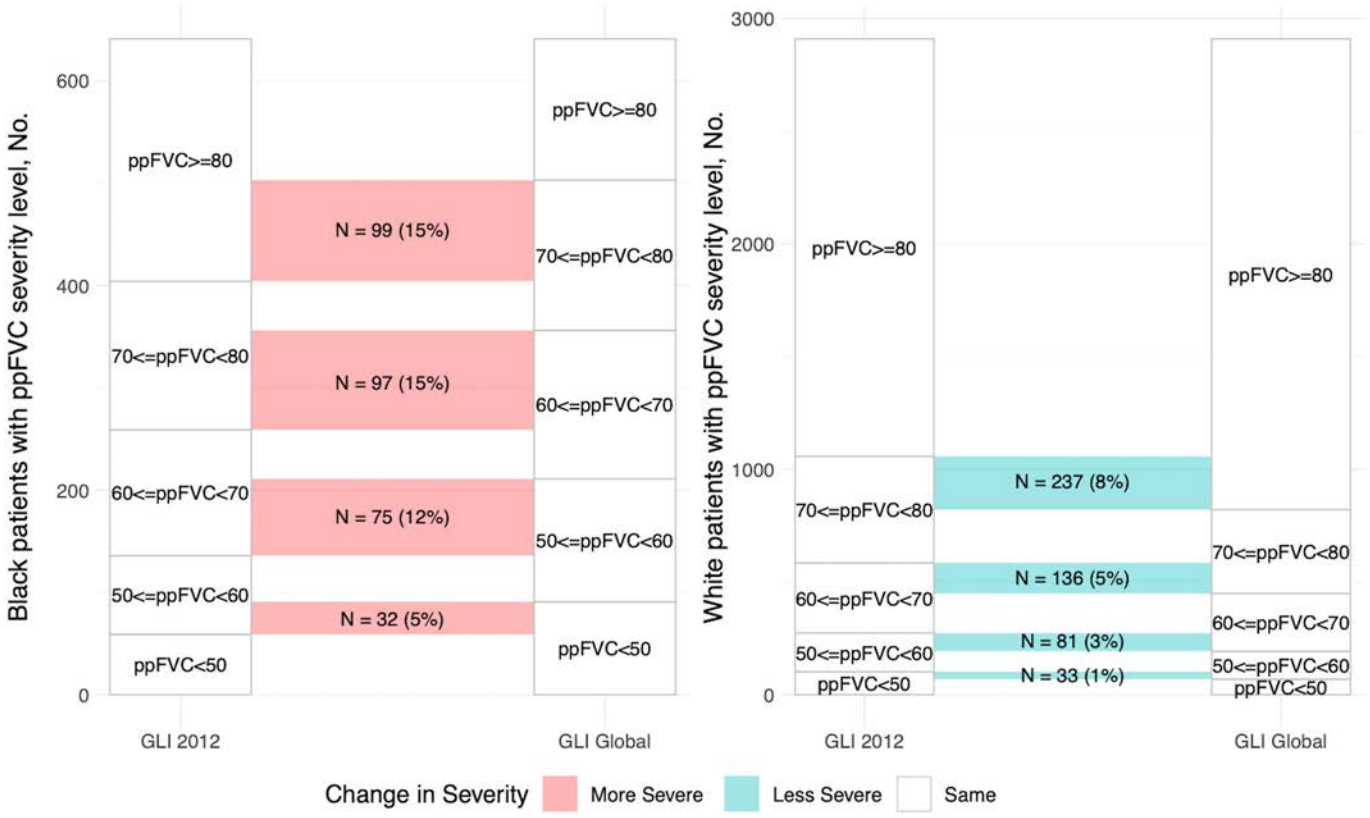

Figure 2. Changes in classification of RLD severity with the GLI 2012 versus GLI Global equations by race. The alluvial plots demonstrate the number and percentage of individuals who switch RLD impairment categories as defined by ppFVC when comparing the race-specific GLI 2012 reference equations and the race-neutral GLI Global equations. Individuals for whom the use of race-neutral equations resulted in more severe RLD are depicted by red alluvia. Individuals for whom the use of race-neutral equations resulted in less severe RLD are depicted by green alluvia. The white bars represent patients who remain in the same RLD category. The height of each alluvium indicates the proportion of individuals switching categories when GLI Global equations are used. GLI, Global Lung Function Initiative; NHANES III, National Health and Nutrition Examination Survey III; ppFVC, percent predicted forced vital capacity; RLD, restrictive lung disease.

Similar results are observed when z scores were used for lung classification. Supplemental Table 1 shows the distributions of individuals in each lung classification category (normal or mild, moderate, and severe RLD) as defined by z score with race-specific versus race-neutral equations. All four Wilcoxon signed rank tests that compared the GLI Global equations to each of the GLI 2012 and NHANES III equations among White and Black individuals were statistically significant ($P < 0.0001$). When the sample was restricted to those who met the 2013 ACR classification criteria, the results were unchanged (data not shown). These data confirm that using race-neutral equations compared to race-specific equations results in significant changes in lung function classification for both races. Supplemental Table 2 shows contingency tables of the same lung classification categories for each race.

DISCUSSION

In this study, we examined whether differing PFT reference equations could impact the classification of RLD in Black and White patients with SSc and influence medical decision-making. Specifically, we evaluated the effect of race-neutral (GLI Global) and race-specific (NHANES III, GLI 2012) equations on the classification of RLD as defined by ppFVC among individuals with SSc through analysis of the JHSC. Our analysis demonstrates that the use of race-specific PFT reference equations leads to clinically significant differences in classification of lung disease among Black and White individuals with SSc. When using the GLI Global equations, 47% of Black individuals were reclassified to more severe lung disease categories as compared to using the GLI 2012 equations. In addition, significantly fewer White individuals would meet clinical ppFVC thresholds for immunosuppression initiation, lung transplant referral, and clinical trial eligibility with the use of race-specific versus race-neutral equations. These data suggest that the historical PFT reference equations may have been a contributing factor to gaps in health outcomes among patients with SSc. Educating clinicians about the variation in reference equations over time, including the transition from race-specific to race-neutral equations, is important to ensure that

Table 2. The potential effect of race-specific vs race-neutral pulmonary function test reference equations on care delivery*

	Eligibility for clinical trials enrollment, n (%)			Initiation of immunosuppression, n (%)			Referral for lung transplant, n (%)		
	NHANES III	GLI 2012	GLI Global	NHANES III	GLI 2012	GLI Global	NHANES III	GLI 2012	GLI Global
All (N = 3,550)	1,449 (41)	1,367 (39)	1,226 (35)	908 (26)	844 (24)	805 (23)	168 (5)	161 (5)	160 (5)
White (n = 2,909)	1,100 (38)	1,000 (34)	783 (27)	660 (23)	585 (20)	449 (15)	115 (4)	102 (4)	69 (2)
Black (n = 641)	349 (54)	367 (57)	443 (69)	248 (39)	259 (40)	356 (56)	53 (8)	59 (9)	91 (14)

* This table demonstrates the numbers and percentages of individuals who meet three clinical criteria with each of the equations (NHANES III, GLI 2012, and GLI Global). ppFVC thresholds for the three clinical criteria were defined as follows: eligibility for clinical trials (SLS I/SLS II), ppFVC < 80% and ppFVC ≥ 45%; initiation of immunosuppression, ppFVC < 70%; referral for lung transplant, ppFVC ≤ 50%. All *P* values from McNemar tests comparing the proportions of individuals who meet each clinical criterion for each race when using race-specific (NHANES III and GLI 2012) and race-neutral (GLI Global) equations were <0.0001. GLI, Global Lung Function Initiative; NHANES III, National Health and Nutrition Examination Survey III; ppFVC, percent predicted forced vital capacity; SLS, Scleroderma Lung Study.

these tests are being interpreted appropriately, both in a research context and when assessing an individual patient's trajectory over time for medical decision-making.

SSc-ILD is a significant cause of morbidity and mortality for Black individuals with SSc. African ancestry is also associated with early pulmonary involvement and increased burden of lung disease.[2,3] Diffuse skin disease, older age at disease onset, shorter disease duration, and anti–topoisomerase 1 antibodies have been identified as risk factors for development or progression of ILD.[3,37] Many studies have attributed the increased rates of ILD among Black individuals to the high prevalence of anti–topoisomerase 1 antibodies. However, other studies suggest that the increased pulmonary disease seen among Black individuals with SSc is multifactorial in nature. Although individuals of all backgrounds may be affected by evolving PFT reference equations, understanding the role measurement tools play in diagnosis, monitoring, and treatment decisions is particularly important in Black patients with SSc given their worse health outcomes.

The incorporation of GLI Global reference equations in the interpretation of PFTs will likely have significant clinical impact for the care of all individuals with SSc. A large cross-sectional study that analyzed all PFTs from an academic center PFT laboratory regardless of underlying disease process found an increased number of Black individuals with RLD, as well as increased severity of the identified impairments, with the use of race-neutral versus race-specific equations.[26] Although our analysis largely focused on the effects of spirometry on Black and White individuals, the implementation of the GLI Global equations will impact individuals of other races as well. A large multicenter study comparing the GLI Global and GLI 2012 equations found that the newer equation resulted in an increase in prevalence of RLD by 51% in Black individuals and 37.1% in Southeast Asian groups.[38]

Further research is needed to understand the impact of varying PFT reference equations on medical decision-making across the broad spectrum of pulmonary diseases. A previous study that evaluated the impact of race-neutral equations on surgeons' decision-making regarding lobectomy for Black individuals with lung cancer found that using race-neutral equations may result in less frequent recommendations for curative lung cancer surgery for Black individuals.[39] This study emphasizes the importance of a multimodal and holistic approach to decision-making for individuals with lung disease. Although the use of race-neutral equations is now recommended by national associations such as the American Thoracic Society, debate exists in the field regarding how to best implement these equations, and it is not clear if race-neutral equations have been universally adopted. A recent longitudinal multicohort study found that race-neutral equations perform similarly to race-specific equations in predicting pulmonary outcomes but affect classification of lung disease. This same study demonstrated that the use of race-neutral equations has significant ramifications on other services, such as occupational eligibility, disability payments, and lung transplant priority.[24]

For rheumatologists, awareness surrounding how PFT reference equations are derived and how these equations evolve over time is necessary to allow for effective clinical decision-making. For example, the recent transition to GLI Global equations may have resulted in shifts in ppFVC for White and Black individuals due to changing equations as opposed to actual clinical change. Without this important context, clinicians may make incorrect interpretations of a patient's disease progression or trajectory. As such, education for rheumatologists regarding these reference equations is essential.

This study evaluated the effect of race-neutral versus race-specific equations on the classification of RLD as defined by ppFVC in a large cohort of Black and White individuals with SSc. Previous studies focused on Black individuals with SSc-ILD have been limited by small sample sizes; by using the JHSC, we were able to compare the use of these equations in a relatively large sample of patients with SSc. However, our study had several limitations. PFTs included in this study were obtained from multiple sites and were not limited to our single center. Given changing practice patterns over more than 30 years, all of our patients did not have high-resolution chest computed tomography or consistent use of right-sided heart catheterization testing to define ILD or PH. Our analysis focuses on individuals who self-identified as Black or White but does not speak to the effect of race-specific or race-neutral equations on individuals of other races or with mixed ethnicities. The American Thoracic Society and European

Thoracic Society now recommend using z scores to classify RLD severity.[15] Although we included additional analysis related to RLD defined by z scores, we framed our analysis on ppFVC because these values are used practically for clinical decision-making. As an exercise to demonstrate how the use of different PFT reference equations could potentially affect care delivery, we used conservative ppFVC thresholds for initiation of immunosuppression and lung transplant referral, recognizing the multifactorial nature of these decisions, including disease trajectory over time. However, this study does not explicitly assess how the use of race-specific versus race-neutral equations affects clinical outcomes for Black and White individuals with SSc.

This study elucidates the impact of race-specific and race-neutral PFT reference equations on classification of RLD among Black and White individuals with SSc. More research is needed to understand how these and other measurement tools may impact the care that individuals with SSc and other systemic autoimmune rheumatic diseases receive. Changing reference equations over time can significantly affect PFT results and interpretation; further education regarding how to appropriately use these tests moving forward is needed.

ACKNOWLEDGMENTS

The authors thank the JHSC patients and staff, including Adrianne Woods, Gwen Leatherman, Cheryl Clare, Michelle Leatherman, and Lauren Fisher.

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 Shah and Kim 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. Gelber AC, Manno RL, Shah AA, et al. Race and association with disease manifestations and mortality in scleroderma: a 20-year experience at the Johns Hopkins Scleroderma Center and review of the literature. Medicine (Baltimore) 2013;92(4):191–205.
2. Moore DF, Kramer E, Eltaraboulsi R, et al. Increased morbidity and mortality of scleroderma in African Americans compared to non-African Americans. Arthritis Care Res (Hoboken) 2019;71(9):1154–1163.
3. McNearney TA, Reveille JD, Fischbach M, et al. Pulmonary involvement in systemic sclerosis: associations with genetic, serologic, sociodemographic, and behavioral factors. Arthritis Rheum 2007;57(2):318–326.
4. Gourh P, Remmers EF, Boyden SE, et al. Brief report: whole-exome sequencing to identify rare variants and gene networks that increase susceptibility to scleroderma in African Americans. Arthritis Rheumatol 2018;70(10):1654–1660.
5. Nietert PJ, Silver RM. Patterns of hospital admissions and emergency room visits among patients with scleroderma in South Carolina, USA. J Rheumatol 2003;30(6):1238–1243.
6. Laing TJ, Gillespie BW, Toth MB, et al. Racial differences in scleroderma among women in Michigan. Arthritis Rheum 1997;40(4):734–742.
7. Volkmann ER, Steen V, Li N, et al. Racial disparities in systemic sclerosis: short- and long-term outcomes among African American participants of SLS I and II. ACR Open Rheumatol 2021;3(1):8–16.
8. Greidinger EL, Flaherty KT, White B, et al. African-American race and antibodies to topoisomerase I are associated with increased severity of scleroderma lung disease. Chest 1998;114(3):801–807.
9. Morgan ND, Shah AA, Mayes MD, et al. Clinical and serological features of systemic sclerosis in a multicenter African American cohort: analysis of the genome research in African American scleroderma patients clinical database. Medicine (Baltimore) 2017;96(51):e8980.
10. Kuchinad KE, Kim JS, Woods A, et al. Racial variability in immune responses only partially explains differential systemic sclerosis disease severity. Ann Rheum Dis 2024;83(11):1513–1521.
11. Johnson SR, Bernstein EJ, Bolster MB, et al. 2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) guideline for the screening and monitoring of interstitial lung disease in people with systemic autoimmune rheumatic diseases. Arthritis Care Res (Hoboken) 2024;76(8):1070–1082.
12. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999;159(1):179–187.
13. Quanjer PH, Stanojevic S, Cole TJ, et al; ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J 2012;40(6):1324–1343.
14. Graham BL, Steenbruggen I, Miller MR, et al. Standardization of Spirometry 2019 update. An official American Thoracic Society and European Respiratory Society technical statement. Am J Respir Crit Care Med 2019;200(8):e70–e88.
15. Stanojevic S, Kaminsky DA, Miller MR, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. Eur Respir J 2022;60(1):2101499.
16. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. Eur Respir J 2005;26(5):948–968.
17. Baugh AD, Shiboski S, Hansel NN, et al. Reconsidering the utility of race-specific lung function prediction equations. Am J Respir Crit Care Med 2022;205(7):819–829.
18. Bhakta NR, Bime C, Kaminsky DA, et al. Race and ethnicity in pulmonary function test interpretation: an official American Thoracic Society Statement. Am J Respir Crit Care Med 2023;207(8):978–995.
19. Elmaleh-Sachs A, Balte P, Oelsner EC, et al. Race/ethnicity, spirometry reference equations, and prediction of incident clinical events: the Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study. Am J Respir Crit Care Med 2022;205(6):700–710.
20. Ekström M, Mannino D. Research race-specific reference values and lung function impairment, breathlessness and prognosis: analysis of NHANES 2007-2012. Respir Res 2022;23(1):271.
21. Gaffney AW, McCormick D, Woolhandler S, et al. Prognostic implications of differences in forced vital capacity in Black and White US adults: findings from NHANES III with long-term mortality follow-up. EClinicalMedicine 2021;39:101073.
22. McCormack MC, Balasubramanian A, Matsui EC, et al. Race, lung function, and long-term mortality in the National Health and Nutrition Examination Survey III. Am J Respir Crit Care Med 2022;205(6):723–724.

23. Regan EA, Lowe ME, Make BJ, et al. Early evidence of chronic obstructive pulmonary disease obscured by race-specific prediction equations. Am J Respir Crit Care Med 2024;209(1):59–69.
24. Diao JA, He Y, Khazanchi R, et al. Implications of race adjustment in lung-function equations. N Engl J Med 2024;390(22):2083–2097.
25. Bowerman C, Bhakta NR, Brazzale D, et al. A race-neutral approach to the interpretation of lung function measurements. Am J Respir Crit Care Med 2023;207(6):768–774.
26. Moffett AT, Bowerman C, Stanojevic S, et al. Global, race-neutral reference equations and pulmonary function test interpretation. JAMA Netw Open 2023;6(6):e2316174.
27. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65(11):2737–2747.
28. Masi AT; Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23(5):581–590.
29. Mukerjee D, St George D, Knight C, et al. Echocardiography and pulmonary function as screening tests for pulmonary arterial hypertension in systemic sclerosis. Rheumatology (Oxford) 2004;43(4):461–466.
30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289–300.
31. Stanojevic S, Graham BL, Cooper BG, et al; Global Lung Function Initiative TLCO working group; Global Lung Function Initiative (GLI) TLCO. Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians. Eur Respir J 2017;50(3):1700010.
32. Tashkin DP, Elashoff R, Clements PJ, et al; Scleroderma Lung Study Research Group. Cyclophosphamide versus placebo in scleroderma lung disease. N Engl J Med 2006;354(25):2655–2666.
33. Tashkin DP, Roth MD, Clements PJ, et al; Scleroderma Lung Study II Investigators. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. Lancet Respir Med 2016;4(9):708–719.
34. Distler O, Assassi S, Cottin V, et al. Predictors of progression in systemic sclerosis patients with interstitial lung disease. Eur Respir J 2020;55(5):1902026.
35. Perelas A, Silver RM, Arrossi AV, et al. Systemic sclerosis-associated interstitial lung disease. Lancet Respir Med 2020;8(3):304–320.
36. Pradère P, Tudorache I, Magnusson J, et al; Working Group on Heart/Lung Transplantation in Systemic Sclerosis. Lung transplantation for scleroderma lung disease: an international, multicenter, observational cohort study. J Heart Lung Transplant 2018;37(7):903–911.
37. Cottin V, Brown KK. Interstitial lung disease associated with systemic sclerosis (SSc-ILD). Respir Res 2019;20(1):13.
38. Kanj AN, Scanlon PD, Yadav H, et al. Application of Global Lung Function Initiative global spirometry reference equations across a large, multicenter pulmonary function lab population. Am J Respir Crit Care Med 2024;209(1):83–90.
39. Bonner SN, Lagisetty K, Reddy RM, et al. Clinical implications of removing race-corrected pulmonary function tests for African American patients requiring surgery for lung cancer. JAMA Surg 2023;158(10):1061–1068.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 478–488
DOI 10.1002/acr.25630

American College of Rheumatology
Empowering Rheumatology Professionals

Physician Referral Patterns to Physical Therapists for Managing Knee Osteoarthritis: A Retrospective Analysis of Electronic Health Records From an Integrated Health System

Samannaaz S. Khoja,[1] Joel M. Stevans,[1] Gustavo J. Almeida,[2] Clair Smith,[1] and Janet K. Freburger[1]

Objective. This study aims to describe the frequency and timing of physician referrals to physical therapists (PTs) and other treatments prescribed over 12 months in patients with recent onset of knee osteoarthritis (KOA). The study also aims to identify determinants of early PT referrals.

Methods. A retrospective study was conducted using electronic medical records from an integrated health care system in the United States. Incident KOA visits between October 2016 and September 2017 were identified using International Classification of Diseases, Tenth Revision codes. Early PT referral was defined as a referral within 15 days of the incident KOA visit by the attending physician. Data on PT referrals, other KOA treatments, patient characteristics, physician specialty, practice location, and volume of patients with KOA were extracted. The number of licensed PTs within the county of the included practices was also obtained. We used generalized linear mixed models to predict factors associated with early PT referrals.

Results. Included in our study were 9,835 patients with incident KOA. Within the 12-month period, 26% of patients had at least one PT referral (n = 2,550) and the median (interquartile range) days to the first referral was 0 (0–34) days. Early PT referral by the index physician occurred in 16.5% (n = 1,629) of patients, was less likely for patients with higher or missing knee pain scores and in rural practices, and more likely in women, patients with higher body mass index, and those seen in counties with a greater supply of licensed PTs.

Conclusion. Physician referrals to PT for a new episode of KOA are infrequent and influenced by patient and practice factors.

INTRODUCTION

Knee osteoarthritis (KOA) is a highly prevalent and costly chronic condition with no known cure. With increasing life expectancy and a growing aging population, the clinical and economic burdens associated with KOA-related treatment and disability are expected to rise significantly in the coming decades.[1] Although total knee arthroplasty (TKA) is the most effective treatment for end-stage KOA, individuals may live several years with chronic knee pain and osteoarthritis-related disability before undergoing TKA.

Since 1995, various clinical practice guidelines have recommended evidence-based nonpharmacologic approaches such as exercise, weight management, and self-management as frontline treatments, alongside nonsteroidal anti-inflammatory drugs (NSAIDs) to manage and reduce pain and disability.[2–6] These guidelines imply support for multidisciplinary teams to manage individuals with KOA, including physical therapists (PTs), who are uniquely qualified to provide exercise, self-management counseling, and targeted treatments (eg, manual therapy, balance and functional exercises) to manage disability. However, the initial management of KOA in the United States is largely driven by

Supported by the Foundation for Physical Therapy's Center on Health Services Training & Research (CoHSTAR), funded by Foundation of Physical Therapy Research and Academy of Orthopaedic Physical Therapy through the American Physical Therapy Association in collaboration with the Academy of Orthopedic Physical Therapy and by the NIH (grant UL1-TR-001857).

[1]Samannaaz S. Khoja, PT, PhD, Joel M. Stevans, DC, PhD, Clair Smith, MS, Janet K. Freburger, PT, PhD: University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Gustavo J. Almeida, PT, PhD: University of Texas Health Science Center at San Antonio.

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.25630).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25630.

Address correspondence via email to Samannaaz S. Khoja, PT, PhD, at ssk21@pitt.edu.

Submitted for publication February 17, 2025; accepted in revised form August 5, 2025.

SIGNIFICANCE & INNOVATIONS

- Despite several clinical practice guidelines recommending nonpharmacologic treatments as frontline approaches for managing knee osteoarthritis (KOA), these interventions remain underused in clinical practice.
- This study provides valuable longitudinal data on the treatments recommended by physicians to patients with incident KOA over a 12-month period, highlighting trends and gaps in care.
- Early physical therapist (PT) referrals were the most common nonpharmacologic approach that occurred in only 16.5% of the cohort identified with KOA. An additional 8% of the cohort with KOA received a PT referral after 15 days from the incident visit by their primary physician or another physician. Oral pain medications and therapeutic injections were more frequently prescribed compared with PT and lifestyle modifications.
- Early PT referrals were influenced by factors such as patient and practice characteristics, underscoring the need to continue addressing opportunities to improve the access and delivery of evidence-based care for patients with KOA.

physicians, with physician referrals being main route through which patients access physical therapy.

Despite updates in clinical practice guidelines, there remains a gap between current practice patterns and guideline-based care. Specifically, physician referral to PTs tends to be suboptimal.[7–9] Given that the guidelines recommend self-management and exercise-based interventions as frontline approaches for all individuals with KOA, we expect that referral to PT and other nonpharmacologic providers would be more common. Our previous work using the National Ambulatory Medical Care Survey showed a decline in PT referrals for KOA-related orthopedic visits from 158 to 88 per 1,000 visits between 2007 to 2009 and 2013 to 2015, whereas referrals for KOA-related primary care physician (PCP) visits remained low but steady, at 26 to 46 per 1,000 visits in the same period.[7] In contrast, pain medication prescriptions for both NSAIDs (132–278 per 1,000 orthopedic visits; 221–498 per 1,000 PCP visits) and narcotics (77–236 per 1,000 orthopedic visits; 233–316 per 1,000 PCP visits) increased significantly within the same timeframe.[7] Another study demonstrated similar findings with referral rates of 16% among all musculoskeletal disorders, including KOA,[10] whereas one study from Australia indicated a lower PT referral rate of 5% for KOA.[9] Additionally, the COVID-19 pandemic has further contributed to a reduction in PT referrals for KOA.[11]

When referrals to PT remain low, the overall receipt of PT will remain low. Dhawan et al reported in a large cohort study using the 2004 to 2009 United Healthcare Database that only 25% of individuals with KOA received PT before TKA.[12] Similarly, an analysis of the University of Utah health care medical records between from 2015 to 2018 showed that only 24% of patients referred to PT for knee pain actually used PT services.[10] These low referral rates and low patient follow-through to PT services coupled with increasing prescription rate for pain medications, including narcotics, suggest that those patients with KOA may not be receiving optimal, guideline-based care.

Early treatment by PTs is associated with favorable patient outcomes and reductions in subsequent narcotics use, health care usage, and costs for knee pain.[13–16] However, current evidence suggests that most patients with KOA may not be adequately counseled about the benefits of PT, let alone receive a referral. Importantly, there is limited longitudinal evidence on the trajectory of care of individuals with incident KOA and factors that may influence the early referral of individuals with KOA to PT providers. It is likely that early referral to PT is not just driven by patient clinical characteristics but also by contextual factors surrounding their care. The Consolidated Framework for Implementation Research (CFIR) emphasizes the importance of examining contextual factors such as the inner settings (eg, clinic characteristics), outer setting (eg, supply of clinicians, insurance policies), and the individuals involved in health care delivery (eg, physicians) that may influence the uptake of evidence-based approaches.[17,18] Hence, an understanding of contextual factors associated with PT referral for KOA at the patient, physician, and practice levels may help inform efforts to improve the use of PT for KOA. The aims of this study were to (1) describe the frequency and timing of physician referral to PT for patients with new-onset KOA diagnosis; (2) describe the frequency of other KOA treatments or procedures prescribed within 12 months from the initial KOA-related visit; and (3) identify patient, physician, and practice-level determinants of early PT referral for KOA.

PATIENTS AND METHODS

Study design and data source. We conducted a retrospective cohort study using electronic medical records (EMRs) of patients seeking ambulatory care within a large, integrated health care system in western Pennsylvania. Limited EMR datasets were obtained from a research data warehouse[19] via an honest broker following the establishment of a data use agreement with the health care system. The study was deemed exempt by the University of Pittsburgh Institutional Review Board.

We identified visits by patients who sought physician care for a new episode of KOA between October 1, 2016 and September 30, 2017 within one of the approximately 350 ambulatory care clinics within the system. A new episode was defined as an ambulatory care visit for KOA preceded by 12 months without any ambulatory care visits for KOA. We then gathered data on all subsequent KOA-related procedures and treatments the patient

received in the 12 months following the index physician visit (ie, between October 1, 2017 and September 30, 2018).

Cohort identification and follow-up. Patient records were included in the retrospective analysis if they (1) had an index visit for a new episode of KOA (defined as a visit between October 1, 2016 and September 30, 2017 related to KOA preceded by a clean period of 12 months with no KOA-related visits), (2) were 45 years of age or older at the time of the index visit, and (3) KOA was recorded as the primary diagnosis, or as a secondary diagnosis with a related KOA procedure recorded at the visit (eg, knee x-ray). The KOA diagnosis was identified using ICD-10 (International Classification of Diseases, Tenth Revision) Clinical Modification codes (Supplementary Table 1). We used codes directly related to KOA (eg, M17X) as well as generic knee pain or swelling codes (eg, M25.56X) because KOA has shown to be underreported in the EMR and is usually coded as generic knee pain.[20] Because we wanted to identify the earliest visit related to KOA, we included individuals with generic knee pain and swelling codes as their primary or secondary diagnosis, provided there were no other non-KOA knee disorders, such as patellofemoral pain (ICD-10 M22) or meniscal or ligament injuries (ICD-10 M23), identified as a primary or secondary diagnosis at that visit. We did not exclude those with unilateral TKA if they were seeking care for their nonoperated knee but excluded those with evidence of bilateral TKA (ICD-10 Z96.653, V43.65) before their initial KOA visit.

Dependent variable. Our primary dependent variable "early referral to PT" was defined as a PT referral provided by the index physician at the index KOA visit or within 15 days of that visit.[13] We also collected data on PT referrals between 16 and 365 days from the index visit. PT referrals were identified via internal custom codes from the orders field in the EMR (Supplementary Table 2).

To describe downstream KOA care over a 12-month period in our cohort, we collected data on referrals to lifestyle or exercise counseling consults, referrals to specialty physicians (ie, orthopedists, physical medicine and rehabilitation, sports medicine), knee-related tests or procedures obtained (ie, therapeutic injections, imaging, surgery), and pain medications prescribed (topical and oral NSAIDs, tramadol, and nontramadol narcotics). Supplementary Tables 2 to 4 lists the downstream treatments and procedures that were collected. These data were collected from the index visit through the 12-month follow-up period for each patient. All downstream health care services associated with a KOA diagnosis for each patient were extracted regardless of the clinician providing the services.

Explanatory variables. Selection of our independent variables of interest was guided by our previous work and reported literature regarding the barriers and facilitators related to health care processes and decisions.[7,21] We hypothesized that factors related to the health care environment such as practice characteristics related to the volume of patients with KOA (inner setting), practice location and supply of licensed PTs (outer settings), individuals involved in providing health care (ie, physician), and recipients of care (ie, patients with KOA) would be associated with the likelihood of receiving a PT referral from a physician. We used the CFIR to inform the selection of our explanatory variables from the EMR (Supplementary Table 5).[17]

Physician characteristics available in the EMR included physician specialty and their volume of patients with KOA seen. The volume of patients with KOA was defined as the total number of identified patients with KOA (within the study period, ie, between October 1, 2016 to September 30, 2018) managed by each physician within our dataset. We hypothesized that physicians with a higher volume of patients with KOA would have different referral patterns or KOA management practices compared with those with a lower volume of patients with KOA. Practice characteristics included urban or rural location based on the designation of the county where the practices were located;[22] availability of PTs within the county where the practice was located, calculated as the number of registered licensed PTs per 10,000 county residents,[23] and volume of patients with KOA at the practice level, calculated as the percentage of patients with KOA relative to the total number of patients seen in the practices during the study data collection period.

Patient characteristics extracted from the EMR included demographic information (ie, age, sex, self-reported race and ethnicity, and insurance) and clinical information (ie, body mass index, smoking status, presence of comorbid conditions, and pain levels). To capture comorbid conditions, we calculated an Elixhauser summary score for each patient based on the associated ICD-10 diagnoses present in their EMR data.[24,25]

Data extraction and statistical analysis. Raw data files with the limited dataset were obtained by the honest broker and then processed to remove duplicates and noneligible encounters (nonphysician visits, no KOA-related diagnoses, encounter outside of the study period, missing information on index physician specialty). Variables of interest were identified via structured data fields such as Current Procedural Terminology codes and internal custom codes (Supplementary Tables 2–4). We reformatted or collapsed categories for variables (eg, insurance type, provider specialty, smoking status) as appropriate. A clean analytic file was then exported to the SAS statistical package for statistical analysis.

We used descriptive statistics to characterize patient, physician, and practice characteristics, to calculate the proportion (%) of patients who received a PT referral (within 15 days of index visit or after) and over 12 months, and to summarize the distribution of time to receipt of first PT referral. Similarly, we used descriptive statistics (frequency and proportions) to describe other KOA

referrals, treatments, and procedures at the index visit and over the 12-month period.

We used generalized linear mixed models (GLMMs) to examine predictors of early referral to PT. The models had a logit link and practice as the random intercept to control for clustered design of the study (patients are clustered by practice). We assessed for multicollinearity among our predictor variables with the phi correlation coefficient. If two variables were highly correlated (ie, coefficient > 0.8) we made decisions to drop one or the other for methodologic (eg, missingness) or conceptual reasons. We then used backward elimination to retain variables that were significant at $P < 0.05$). For observations with missing data among the predictor variables, we created an additional "missing" category for those predictors and included them in the model. We conducted one sensitivity analysis to examine whether determinants of early PT referral differed among those with ICD-10 codes specific to KOA compared with those with only general knee symptoms codes. Analyses were done using SAS (version 9.4) software.

RESULTS

We identified 10,101 eligible patients with a primary or secondary diagnosis related to KOA from a total of 27,160 patient records with an ambulatory care visit performed by a physician during the study period. Patient records were excluded if they were duplicative (ie, for a patient with more than one eligible visit only the earliest visit was retained), missing provider specialty information, age less than 45 years at index visit, visit date beyond data collection period, or had a noneligible knee diagnosis. Additionally, patients were excluded from the final analytic dataset if the practice location was unknown or missing, or if they were seen in practices with five or fewer total patients with KOA. Because practice was determined to be a random effect, we excluded practices with five or fewer observations due to model convergence issues. See Supplementary Figure 1 for list of cohort exclusions. The final analytic dataset included 9,835 patients with KOA.

Overall, the sample was older than 60 years of age, mostly women (62%), non-Hispanic White (87%), and had an average body mass index (BMI) in the obese category (>30). Most patients were seen by a PCP or orthopedist at the index visit. Table 1 describes sample characteristics based on whether early PT referral was received or not. Missing data were observed for pain, BMI, and volume of patients with KOA per physician and per practice. For the practice-level volume of patients with KOA, the missingness was less than 1%; thus, we did not create a separate category. For the other three variables (ie, pain, BMI, and volume of patients with KOA per physician), we created a "missing" category and included it in the model to account for the incomplete data.

Of the 9,835 patients included in this analysis, 16.5% (n = 1,629) received an early PT referral from their index physician. Orthopedists had the largest proportion of patients receiving early PT referral (Table 1 and Figure 1A), and most PT referrals (n = 1,513) were provided at the index visit (Figure 1B). A small proportion of the cohort received PT referrals within 15 days (1.3%, n = 127) or beyond 15 days (5.1%, n = 502) from that initial visit from providers other than their index physician. Additionally, a small proportion of the cohort received a PT referral from their index physician between 16 and 365 days from their initial visit (~3%, n = 293). The total number of PT referrals per patient within the 12-month period in the cohort with KOA ranged from zero to six referrals; around 73% of patients had zero PT referrals, 21% had at least one PT referral, and 5% had two or more PT referrals. The median (interquartile range) number of days to the first PT referral by any physician after the index visit was 0 (0–34) days. The rate of early PT referral was also similar when looking at patients based on diagnostic category. About 38% of patients identified had a KOA-specific diagnosis, whereas 62% had only general knee symptoms. Early PT referral was observed in 16.7% of patients with a KOA-specific diagnosis and in 16.3% of patients with only general knee symptoms (Supplementary Table 6).

Other referrals, procedures and medications prescribed during the 12-month follow-up are reported in Figures 2 and 3. Around 20% of patients with KOA received referrals to other medical or surgical physician specialties. The most common procedure was knee x-ray (~70%), followed by therapeutic injections (~26%). The most common type of medications prescribed were NSAIDs (~41%) followed by nontramadol narcotics (~25%).

The multicollinearity analysis revealed only one significant relationship among the predictor variables: physician specialty and physician volume of patients with KOA. We retained physician specialty in the model because it had no missing values, whereas physician volume of patients with KOA had a small percentage (<1%) of missing values. The results of the GLMM analysis are presented in Tables 2 and 3. Among patient characteristics, women were more likely to receive an early PT referral, whereas patients with missing or higher pain scores were less likely to receive an early PT referral. BMI as a category was significantly associated with early PT referral; however, the odds ratios of individual categories (overweight, obese, or missing) were not significantly different from the reference group (BMI < 25). Physician specialty was not significantly associated with the likelihood of receiving an early PT referral. Among practice characteristics, location and supply of PTs were significantly associated with the receipt of PT referral. Patients seen in rural practices were less likely to receive an early PT referral, and those seen in counties with greater PT availability were more likely to receive an early PT referral. Sensitivity analysis based on KOA diagnostic category did not reveal any major differences in the magnitude or direction of associations between our dependent variable and

Table 1. Characteristics of the cohort with KOA*

Characteristics	Early PT referral (n = 1,629)	PT referral beyond 15 days or no referral (n = 8,206)
Patient characteristics		
Age, mean ± SD, y	63.7 ± 10.5	65.1 ± 11.1
Age category, n (%), y		
45–59	590 (36)	2,670 (33)
59–69	562 (34)	2,779 (34)
>69	477 (29)	2,757 (34)
BMI, mean ± SD	31.9 ± 7.4	31.9 ± 7.3
BMI category, n (%)		
<25	243 (15)	1,119 (14)
25–30	435 (27)	2,199 (27)
>30	805 (49)	4,005 (49)
Missing	146 (9)	883 (11)
Female sex, n (%)	1,096 (67)	4,960 (60)
Race, n (%)		
White	1,392 (85)	7,412 (90)
Black or African American	186 (11)	637 (8)
Other, declined, or unknown[a]	51 (3)	157 (2)
Ethnicity, n (%)		
Hispanic or Latino	9 (1)	55 (1)
Not Hispanic or Latino	1,558 (96)	7,933 (97)
Declined or unknown	62 (4)	218 (3)
Insurance type, n (%)		
Commercial or private	960 (59)	4,680 (57)
Medicaid	140 (9)	635 (8)
Medicare	503 (31)	2,783 (34)
Other	26 (2)	108 (1)
Smoking status, n (%)		
Never smoker	893 (55)	4,262 (52)
Current smoker	172 (11)	944 (12)
Former smoker	561 (3)	2,987 (36)
Unknown	3 (0)	13 (0)
Pain, mean ± SD, numeric pain scale (0–10)	5.5 ± 2.6	6 ± 2.5
Pain category, n (%)		
0–3	165 (10)	439 (5)
4–6	255 (16)	942 (11)
7–10	275 (17)	1,102 (13)
Missing pain scores	934 (57)	5,723 (70)
Practice characteristics		
Rural practice location, n (%)	312 (19)	2,671 (33)
County-level number of PTs per 10,000 residents, mean ± SD	11.4 ± 1.8	10.8 ± 2.2
Number of PTs per 10,000 residents, n (%)		
<10 PTs per 10,000 residents	215 (13)	2,031 (25)
>10 PTs per 10,000 residents	1,414 (87)	6,175 (75)
Physician characteristics		
Physician specialty, n (%)		
Orthopedic surgeon	892 (55)	3,168 (39)
Primary care physician	610 (37)	4,340 (53)
Other	127 (8)	698 (9)
Volume of patients with KOA, n (%)		
0–14	505 (31)	2,986 (36)
15–79	395 (24)	2,693 (33)
>80	725 (45)	2,495 (30)
Missing	4 (0)	32 (0)
Patients with KOA as a proportion to total practice volume, n (%)		
≤2.3	190 (12)	1,455 (18)
2.4–3.7	236 (14)	1,586 (19)
≥3.8	230 (14)	1,362 (17)
Missing	973 (60)	3,803 (46)
Elixhauser Comorbidity index, n (%)		
0 or 1 comorbidity	494 (30)	2,341 (29)

(Continued)

Table 1. (*Cont'd*)

Characteristics	Early PT referral (n = 1,629)	PT referral beyond 15 days or no referral (n = 8,206)
2 or 3 comorbidities	529 (32)	2,669 (33)
4 or 5 comorbidities	338 (21)	1,675 (20)
≥6 comorbidities	268 (16)	1,521 (19)

* BMI, body mass index; EMR, electronic medical record; KOA, knee osteoarthritis; PT, physical therapist.
[a] Includes race categories of American Indian or Alaskan Native, Native Hawaiian or other pacific islander, Asian, multiracial origin, and if declined or not reported in the EMR.

the predictor variables when compared to the full model (Supplementary Tables 7 and 8). Briefly, sensitivity analysis did reveal a stronger magnitude of association between PT availability and location with early PT referral in the KOA-specific diagnosis subcohort, whereas association between BMI and early PT referral was stronger in the general knee symptom subcohort.

DISCUSSION

Our study describes the care that patients with incident KOA received over a 12-month period in a large integrated health system. It also examines factors at the patient, practice, and physician levels that are associated with receipt of early PT referrals. We found a low rate of PT referrals for KOA (16.5%), consistent with previous reports that have reported PT referral rates for KOA between 5 and 16%.[9,10,12] Early PT referrals were most common when the index physicians were orthopedic surgeons compared with PCPs or other specialists. Some possible explanations include orthopedic surgeons having greater experience with rehabilitation approaches and guideline-based care for KOA, the patient not being a candidate for surgery, or the patient needing conservative management before considering surgery.

A unique finding from our study pertains to the timing of PT referrals. Our results indicate that over 50% of PT referrals occurred at the index visit and up to 75% of them occurred within 34 days of that index visit. Further, Figure 1B shows a sharp decline in PT referrals around two weeks from the index visit. Among patients who received a PT referral, most were referred only once, either by the index physician or by another physician within the 12-month follow-up period. These observations indicate that PT referrals tend to occur early in the management of KOA and are rarely reconsidered later in the 12-month care trajectory. This pattern highlights a potential missed opportunity to incorporate PT as a core component of ongoing conservative care for KOA. Considering PT early on during the management of KOA is critical due to its impact on patient outcomes and downstream health care use. Previous studies have indicated that early PT initiation has implications for lowering risk related to the use of opioids,[13,15] intra-articular injections,[13,26] and knee surgery.[13,27]

Our findings align with previous studies that have shown that nonpharmacologic approaches are underused relative to pharmacological approaches for KOA.[9,28] PT referral was the most frequent nonpharmacologic approach (16.5% of patients), whereas lifestyle modifications were documented in only 0.3% of patients and referrals for complementary and alternative medicine were nearly nonexistent in our dataset. Although not all patients with KOA require a PT referral, we would expect that most would be educated about exercise, lifestyle, and self-management interventions because they are strongly recommended in recent

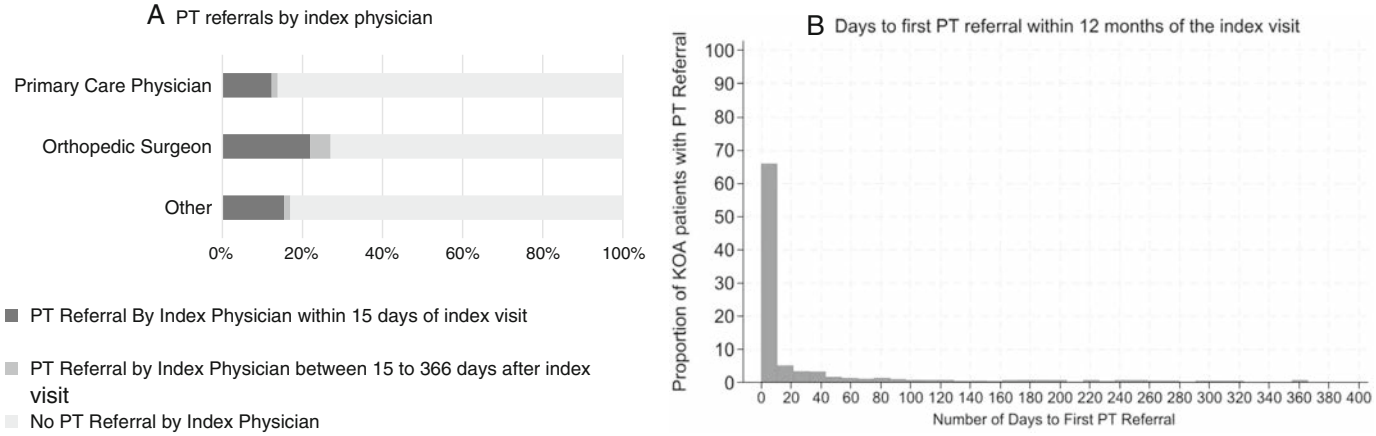

Figure 1. Rate and timing of the first PT referral received in the cohort with KOA over the 12-month follow-up period. KOA, knee osteoarthritis; PT, physical therapist.

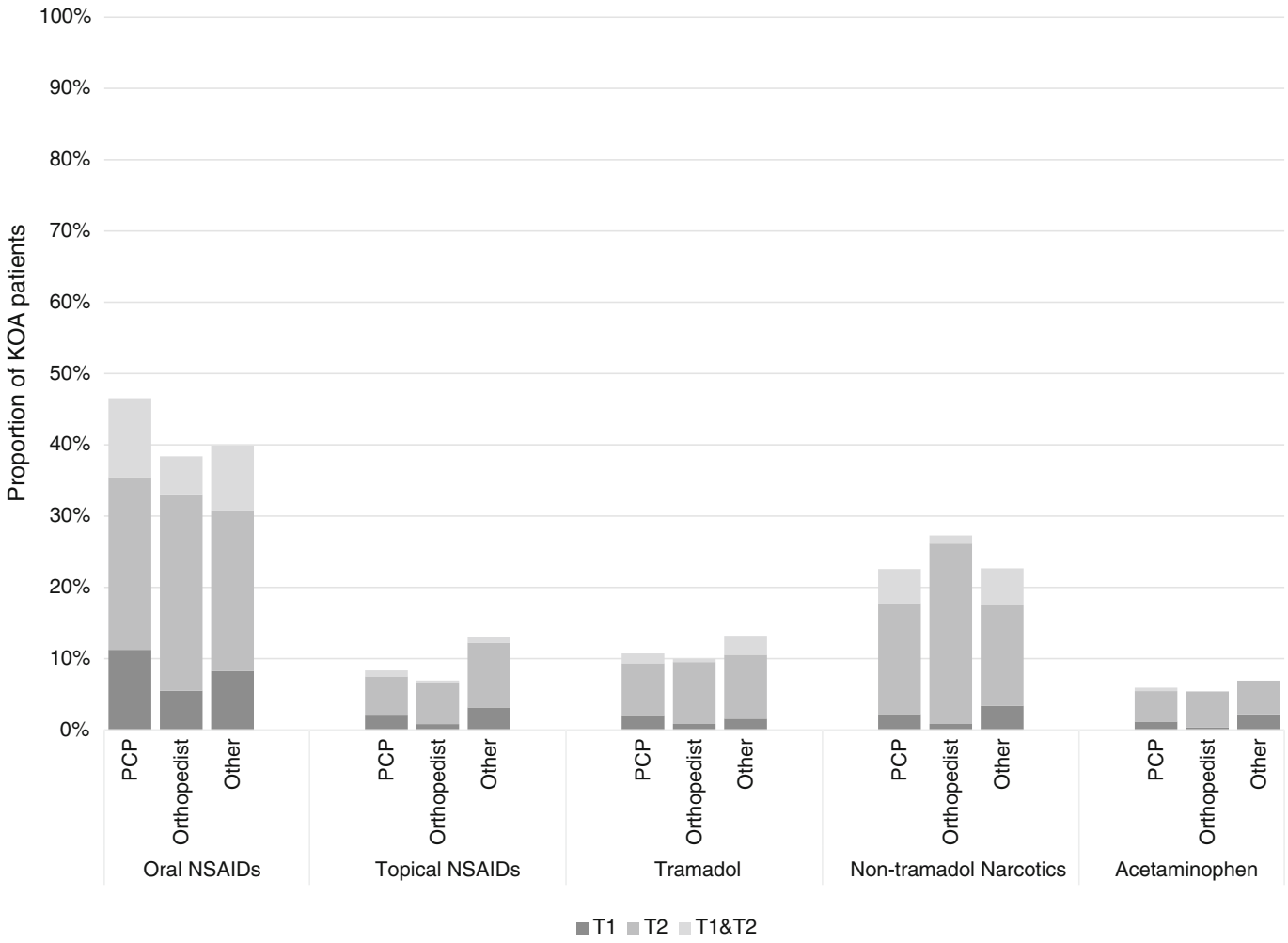

Figure 2. Proportion of patients who were prescribed medications for KOA at index visit (T1), or at a follow-up visit within 12 months of the index visit (T2), or at both index and follow-up visits (T1 and T2). KOA, knee osteoarthritis; NSAIDs, nonsteroidal anti-inflammatory drugs; PCP, primary care physician.

clinical practice guidelines.[2,6] The infrequent use of lifestyle modifications and complementary and alternative medicine in our data could be multifactorial but not limited to physician knowledge, lack of defined referral resources, and inadequate EMR documentation. Relative to nonpharmacologic approaches, use of pain medications (NSAIDs, 41% of patients; nontramadol narcotics, 25% of patients) including therapeutic injections (26% of patients) was higher. Although use of NSAIDs is guideline-concordant, the use of nontramadol narcotics during early KOA management is concerning because the most recent American College of Rheumatology and Arthritis Foundation guidelines conditionally recommend against their use.[2] Therapeutic injections may be appropriate; however, past evidence suggests that therapeutic injections do not have superior benefits over exercise or PT interventions.[14,29] A recent meta-analysis also indicated a potential superior benefit of exercise over NSAIDs, although there is a paucity of trials directly comparing efficacy of pharmacologic versus nonpharmacologic approaches for patients with KOA.[30] Despite their comparable or superior benefits relative to pharmacological approaches, PT or exercise is underprescribed as a frontline approach for patients with KOA.

The final multivariable model identified three patient characteristics and two practice characteristics that were significantly associated with early PT referrals. Among patient characteristics, the reduced likelihood of PT referrals in patients with higher knee pain scores may reflect the preference of physicians and patients for interventions that can be provided immediately at that same visit such as medications or injections to address symptoms before considering PT, or could be driven by missing data and bias in data collection of pain scores. Women were more likely to receive an early PT referral, and this finding aligns with prior research that suggests that women are more likely to use physical therapy.[31,32] With respect to BMI, although no individual categories were significantly different from the referent category, the odds ratios suggest a general trend for individuals with greater BMI to have a higher likelihood of receiving an early PT referral.

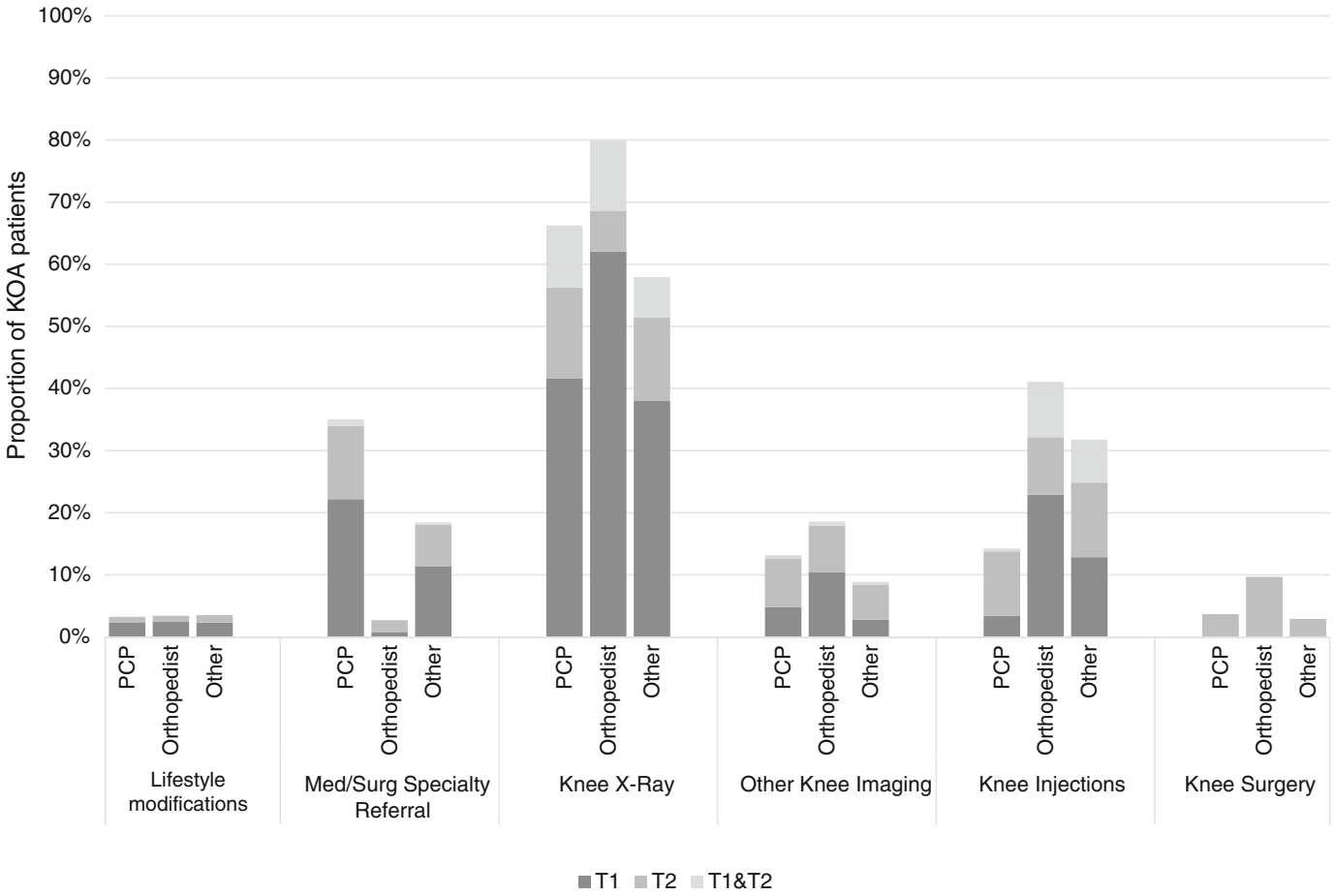

Figure 3. Proportion of patients who received treatments or procedures for KOA at index visit (T1), or at a follow-up visit within 12 months of the index visit (T2), or at both index and follow-up visits (T1 and T2). KOA, knee osteoarthritis; Med/Surg, medical or surgical; PCP, primary care physician.

This may reflect the physician's perception of PT being a higher priority for individuals with higher BMI due to the known association of obesity with KOA symptoms, the overall benefits of exercise for obesity, and potential need to reduce weight before TKA.

In terms of practice characteristics, patients seen in rural practices were less likely to receive early PT referrals compared with those seen in urban practices, and those seen in counties with more than 10 licensed PTs per 10,000 residents were more likely to receive an early referral. These findings reflect geographic disparities in care due to decreased accessibility of resources including supply of PTs. Reduced accessibility to PT services in rural areas has been reported in the literature[7] and is acknowledged as an important factor contributing to health care disparities. Geographic neighborhood-level deprivation has been associated with worse KOA symptoms and outcomes, especially after TKA.[33,34] One recent publication in the French health care system reported lower likelihood PT referral by physicians for musculoskeletal disorders in areas with higher deprivation index.[35] To date, the relationship between early PT referral for patients with KOA and geographic or social deprivation has not been extensively studied and future studies on this topic are warranted to minimize disparities in care and improve accessibility for those with KOA.

The findings of this study suggest that the referral to PT for conservative care of patients with KOA is relatively low and is influenced by factors at the patient, physician, and practice levels. Although we did not assess whether patients who received a PT referral attended PT; we anticipate that this proportion would be even lower, indicating that patients who may benefit from PT are not receiving PT. The study also illustrates the importance of continued investigation and investment in implementing alternative models of care to improve access and use of PT, increase patients' engagement in nonpharmacologic interventions (eg, exercise and lifestyle changes), and improve physician adherence to KOA guideline recommendations. Some examples of KOA models of care being examined to improve access and use of KOA guideline-recommended treatments include telerehabilitation[36–40] and interdisciplinary osteoarthritis programs that include a PT referral or visit defined within the care pathway.[41]

Table 2. Full model showing univariate associations between patient, physician, and practice factors and early PT referral*

Independent variables	OR (95% CI)	*P* value
Patient characteristics		
Age, y		
45–59 (ref)	1.00	0.136
59–69	0.93 (0.80–1.07)	
>69	0.84 (0.71–1.00)	
Sex		
Male (ref)	1.00	**<0.0001**
Female	1.43 (1.27–1.62)	
Race		
White (ref)	1.00	0.221
Black	1.11 (0.88–1.38)	
Other	1.32 (0.92–1.89)	
Ethnicity		
Not Hispanic or Latino (ref)	1.00	0.349
Hispanic or Latino	0.66 (0.31–1.41)	
Declined or missing	1.17 (0.85–1.60)	
Health insurance coverage		
Commercial (ref)	1.00	0.256
Medicaid	1.07 (0.85–1.35)	
Medicare	1.13 (0.98–1.30)	
Other	1.32 (0.82–2.12)	
BMI		
<25 (ref)	1.00	**0.024**
25–30	1.06 (0.88–1.28)	
>30	1.13 (0.95–1.35)	
Missing	0.78 (0.60–1.03)	
Smoking status		
Current smoker (ref)	1.00	0.820
Former smoker	1.05 (0.85–1.29)	
Never smoker	1.09 (0.89–1.33)	
Unknown	1.42 (0.35–5.78)	
Pain category, numeric pain scale (0–10)		
0–3 (ref)	1.00	**0.0004**
4–6	0.76 (0.59–0.97)	
7–10	0.71 (0.55–0.91)	
Missing	0.60 (0.47–0.76)	
Elixhauser Comorbidity index		
0 or 1 comorbidity (ref)	1.00	0.306
2 or 3 comorbidities	0.95 (0.82–1.11)	
4 or 5 comorbidities	0.96 (0.81–1.15)	
≥6 comorbidities	0.84 (0.69–1.01)	
Physician characteristics		
Specialty		
PCP (ref)	1.00	0.683
Orthopedics	1.09 (0.69–1.72)	
Other	1.17 (0.82–1.67)	
Practice characteristics		
PT availability		
<10 PTs per 10,000 residents (ref)	1.00	0.081
>10 PTs per 10,000 residents	1.39 (0.96–2.02)	
Location		
Urban (ref)	1.00	**0.013**
Rural	0.64 (0.46–0.91)	

* Bold values denote significance at $P < 0.05$. BMI, body mass index; CI, confidence interval; OR, odds ratio; PCP, primary care physician; PT, physical therapist; ref, reference group.

Our study is not without limitations. Because we had an observational study using EMR data, it is subject to inherent errors and missing data commonly associated with EMR documentation. There is a possibility that we may have misclassified patients with KOA because we included ICD-10 codes for patients with generic knee pain or symptoms without any other ICD-10 codes indicating alternative pathologies and those with evidence of unilateral TKA seeking care for their nonoperated

Table 3. Final model with multivariable associations between patient, physician, and practice factors and early PT referral*

Independent variables	OR (95% CI)	*P* value
Patient characteristics		
Sex		
Male	1.00	<0.0001
Female	1.43 (1.26–1.61)	
BMI		
<25	1.00	
25–30	1.06 (0.88–1.27)	0.019
>30	1.13 (0.95–1.34)	
Missing	0.78 (0.59–1.02)	
Pain category, numeric pain scale (0–10)		
0–3	1.00	0.0002
4–6	0.76 (0.59–0.97)	
7–10	0.70 (0.54–0.89)	
Missing	0.59 (0.46–0.75)	
Practice characteristics		
PT availability		
<10 PTs per 10,000 residents	1.00	0.048
>10 PTs per 10,000 residents	1.43 (1.00–2.04)	
Location		
Urban	1.00	
Rural	0.65 (0.46–0.91)	0.013

* BMI, body mass index; CI, confidence interval; OR, odds ratio; PT, physical therapist.

knee. Although individuals with unilateral TKA are likely to have a different trajectory of care due to their experience with KOA compared with those without any knee replacement surgery, the number of such patients in our sample were quite small (∼2%) and likely not to impact the overall findings. We acknowledge that some of our findings may be due to sampling bias, missing data, or unique characteristics of patients within our health system compared with the general US population with KOA. The sampled cohort with KOA, although representative of patients in our health care system, was not very diverse (non-White non-Black racial and ethnic groups were < 2%) and does not reflect the population with KOA in the United States. Additionally, only EMR data were available for this study, which limited our ability to examine other recommended CFIR contextual factors such as the influence of patient resources (eg, copay burden, time constraints) and preferences or perceptions about PT (previous PT experience, misconceptions about exercise and osteoarthritis), and physician perceptions about benefits of PT for patients with KOA. This study is also focused on physician behavior (ie, providing a PT referral) and does not account for whether the patient attended a PT session.

Despite these limitations, our study has some notable strengths. It includes a large, representative sample of patients from a major health care system in western Pennsylvania. Also, we captured longitudinal data to assess how patients were managed for one year from their first identifiable visit related to a new episode of care for KOA symptoms.

This study demonstrated that the rate of physician referrals to PT for a new episode of KOA is low, and when present, typically occurs at the incident visit. Early physician referral to PT was influenced by few patient factors (female sex, level of pain, BMI) and by practice factors (urban or rural location). More commonly used KOA treatments over 12-month period from the index visit included therapeutic injections or pharmacological agents.

ACKNOWLEDGMENTS

Authors would like to acknowledge the Department of Health Informatics at the School of Health and Rehabilitation Science, University of Pittsburgh and their Master of Information Science graduate student Jing Pang for support with data management and cleaning. Additionally, the authors acknowledge the limited use of a large language model (ChatGPT) to improve readability and grammar in some sentences during the preparation of this article. Authors reviewed the final content of the article for accuracy and clarity before submission.

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 Khoja 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. 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.
2. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020;72(2):220–233.
3. Brown GA. AAOS clinical practice guideline: treatment of osteoarthritis of the knee: Evidence-based guideline, 2nd edition. J Am Acad Orthop Surg 2013;21(9):577–579.
4. Fernandes L, Hagen KB, Bijlsma JW, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013;72(7):1125–1135.
5. McAlindon TE, Bannuru RR, Sullivan MC, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014;22(3):363–388.
6. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27(11):1578–1589.
7. Khoja SS, Almeida GJ, Freburger JK. Recommendation rates for physical therapy, lifestyle counseling, and pain medications for managing knee osteoarthritis in ambulatory care settings: a cross-sectional analysis of the national ambulatory care survey (2007–2015). Arthritis Care Res (Hoboken) 2020;72(2):184–192.

8. Freburger JK, Khoja S, Carey TS. Primary care physician referral to physical therapy for musculoskeletal conditions, 2003–2014. J Gen Intern Med 2018;33(6):801–803.
9. Bennell KL, Bayram C, Harrison C, et al. Trends in management of hip and knee osteoarthritis in general practice in Australia over an 11-year window: a nationwide cross-sectional survey. Lancet Reg Health West Pac 2021;12:100187.
10. Sharpe JA, Martin BI, Fritz JM, et al. Identifying patients who access musculoskeletal physical therapy: a retrospective cohort analysis. Fam Pract 2021;38(3):203–209.
11. Kaur M, Black D, Fine J, et al. Referrals for physical therapy for osteoarthritis during the COVID-19 pandemic: a retrospective analysis. PLoS One 2021;16(11):e0259679.
12. Dhawan A, Mather RC 3rd, Karas V, et al. An epidemiologic analysis of clinical practice guidelines for non-arthroplasty treatment of osteoarthritis of the knee. Arthroscopy 2014;30(1):65–71.
13. Stevans JM, Fitzgerald GK, Piva SR, et al. Association of early outpatient rehabilitation with health service utilization in managing Medicare beneficiaries with nontraumatic knee pain: retrospective cohort study. Phys Ther 2017;97(6):615–624.
14. Rhon DI, Kim M, Asche CV, et al. Cost-effectiveness of physical therapy vs intra-articular glucocorticoid injection for knee osteoarthritis: a secondary analysis from a randomized clinical trial. JAMA Netw Open 2022;5(1):e2142709.
15. Kumar D, Neogi T, Peloquin C, et al. Delayed timing of physical therapy initiation increases the risk of future opioid use in individuals with knee osteoarthritis: a real-world cohort study. Br J Sports Med 2023;57(15):958–964.
16. Mazzei DR, Ademola A, Abbott JH, et al. Are education, exercise and diet interventions a cost-effective treatment to manage hip and knee osteoarthritis? A systematic review. Osteoarthritis Cartilage 2021; 29(4):456–470.
17. Damschroder LJ, Aron DC, Keith RE, et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50.
18. Damschroder LJ, Reardon CM, Widerquist MAO, et al. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022;17(1):75.
19. Visweswaran S, McLay B, Cappella N, et al. An atomic approach to the design and implementation of a research data warehouse. J Am Med Inform Assoc 2022;29(4):601–608.
20. Yu D, Jordan KP, Peat G. Underrecording of osteoarthritis in United Kingdom primary care electronic health record data. Clin Epidemiol 2018;10:1195–1201.
21. Khoja SS, Almeida GJ, Freburger JK. Recommendation rates for physical therapy, lifestyle counseling, and pain medications for managing knee osteoarthritis in ambulatory care settings: a cross-sectional analysis of the national ambulatory care survey (2007–2015). Arthritis Care Res (Hoboken) 2020; 72(2): 184–192.
22. Rural - urban county map. Center for Rural Pennsylvania website. Accessed January 29, 2025. https://center-for-rural-pennsylvania-data-hub-ruralpa.hub.arcgis.com/pages/county-profiles
23. Welcome to the future of healthcare workforce analysis. National Center for the Analysis of Healthcare Mapping website. Accessed January 29, 2025. http://portals.ncahd.org
24. Linden A. *Elix22: Stata module to compute Elixhauser comorbidity measures*. Version 2022.1. StataCorp. Accessed May 28, 2025. 2022. http://fmwww.bc.edu/RePEc/bocode/e
25. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Med Care 1998;36(1):8–27.
26. Neogi T, Dubreuil M, Peloquin C, et al. Association of physical therapy care with use of intra-articular injections in people with knee osteoarthritis: a real-world cohort study. Arthritis Rheumatol 2025;77(8): 1006–1014.
27. Chen WH, Tsai WC, Wang HT, et al. Can early rehabilitation after osteoarthritis reduce knee and hip arthroplasty risk?: A national representative cohort study. Medicine (Baltimore) 2019;98(21):e15723.
28. Cronström A, Nero H, Lohmander LS, et al. On the waiting list for joint replacement for knee osteoarthritis: are first-line treatment recommendations implemented? Osteoarthr Cartil Open 2020;2(2):100056.
29. Liles SC, Bley B, White DK. The effects of exercise and intra-articular injections versus exercise alone for the treatment of knee osteoarthritis: a scoping review of the evidence. Osteoarthr Cartil Open 2024; 6(2):100456.
30. Thorlund JB, Simic M, Pihl K, et al. Similar effects of exercise therapy, nonsteroidal anti-inflammatory drugs, and opioids for knee osteoarthritis pain: a systematic review with network meta-analysis. J Orthop Sports Phys Ther 2022;52(4):207–216.
31. Bawa HS, Weick JW, Dirschl DR. Gender disparities in osteoarthritis-related health care utilization before total knee arthroplasty. J Arthroplasty 2016;31(10):2115–2118.
32. Iversen MD, Schwartz TA, von Heideken J, et al. Sociodemographic and clinical correlates of physical therapy utilization in adults with symptomatic knee osteoarthritis. Phys Ther 2018;98(8):670–678.
33. Kamath CC, O'Byrne TJ, Lewallen DG, et al. Neighborhood-level socioeconomic deprivation, rurality, and long-term outcomes of patients undergoing total joint arthroplasty: analysis from a large, tertiary care hospital. Mayo Clin Proc Innov Qual Outcomes 2022;6(4): 337–346.
34. Peoples J, Tanner JJ, Bartley EJ, et al. Association of neighborhood-level disadvantage beyond individual sociodemographic factors in patients with or at risk of knee osteoarthritis. BMC Musculoskelet Disord 2024;25(1):887.
35. Peurois M, Bertin M, Fouquet N, et al. Factors associated with referral to physiotherapists for adult patients consulting for musculoskeletal disorders in primary care; an ancillary study to ECOGEN. BMC Prim Care 2023;24(1):13.
36. Allen KD, Arbeeva L, Callahan LF, et al. Physical therapy vs internet-based exercise training for patients with knee osteoarthritis: results of a randomized controlled trial. Osteoarthritis Cartilage 2018;26(3): 383–396.
37. Bennell KL, Campbell PK, Egerton T, et al. Telephone coaching to enhance a home-based physical activity program for knee osteoarthritis: a randomized clinical trial. Arthritis Care Res (Hoboken) 2017; 69(1):84–94.
38. Latif-Zade T, Tucci B, Verbovetskaya D, et al. Systematic review shows tele-rehabilitation might achieve comparable results to office-based rehabilitation for decreasing pain in patients with knee osteoarthritis. Medicina (Kaunas) 2021;57(8):764.
39. Wang HN, Luo P, Liu S, et al. Effectiveness of internet-based telehealth programs in patients with hip or knee osteoarthritis: systematic review and meta-analysis. J Med Internet Res 2024;26:e55576.
40. White DK, Hinman RS, Liles S, et al. A telehealth physical therapy intervention to increase physical activity in adults with knee OA: The Delaware PEAK randomized controlled trial. J Orthop Sports Phys Ther 2025;55(5):377–385.
41. Cunningham J, Doyle F, Ryan JM, et al. Primary care-based models of care for osteoarthritis; a scoping review. Semin Arthritis Rheum 2023; 61:152221.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 489–500
DOI 10.1002/acr.25629

American College of Rheumatology
Empowering Rheumatology Professionals

Hip Morphology–Based Osteoarthritis Risk Prediction Models: Development and External Validation Using Individual Participant Data From the World COACH Consortium

Myrthe A. van den Berg,[1] Fleur Boel,[1] Michiel M. A. van Buuren,[1] Noortje S. Riedstra,[1] Jinchi Tang,[1] Harbeer Ahedi,[2] Nigel K. Arden,[3] Sita M. A. Bierma-Zeinstra,[1] Cindy G. Boer,[1] Flavia M. Cicuttini,[4] Timothy F. Cootes,[5] David T. Felson,[6] Willem Paul Gielis,[7] Joshua Heerey,[8] Graeme Jones,[9] Stefan Kluzek,[10] Nancy E. Lane,[11] Claudia Lindner,[5] Joyce B. J. van Meurs,[1] Andrea Mosler,[8] Amanda E. Nelson,[12] Michael C. Nevitt,[13] Edwin H. Oei,[1] Jos Runhaar,[1] Harrie Weinans,[14] Jesse H. Krijthe,[15] and Rintje Agricola[1]

Objective. This study aims to develop hip morphology-based radiographic hip osteoarthritis (RHOA) risk prediction models and investigates the added predictive value of hip morphology measurements and the generalizability to different populations.

Methods. We combined data from nine prospective cohort studies participating in the Worldwide Collaboration on OsteoArthritis prediCtion for the Hip (World COACH) consortium. RHOA grades were harmonized, and incident RHOA was defined as hips without definite RHOA at baseline that developed definite RHOA within four to eight years. Baseline hip morphology was quantified with automatically and uniformly determined lateral center edge angle and alpha angle measurements on anteroposterior radiographs. Discriminative performance of generalized linear mixed model (GLMM) definitions with and without hip morphology measurements was determined with stratified cross-validation. With leave-one-cohort-out cross-validation, the generalizability to unseen populations of hip morphology–based GLMMs and random forest (RF) models was evaluated.

Results. From the included 35,984 hips without definite RHOA at baseline, 4.7% developed incident RHOA within four to eight years. The GLMM with cohort-specific intercept, considering baseline demographics, RHOA grade, and hip morphology measurements, showed a mean area under the receiver operating characteristic curve (AUC) of 0.80 (±0.01) in stratified cross-validation. Using a marginal intercept decreased performance by 0.1 in AUC. Similar results were found for a GLMM without hip morphology measurements. Leave-one-cohort-out cross-validation showed comparable discrimination (AUC between 0.56–0.88) and calibration performance for hip morphology-based GLMMs and RF models.

Conclusion. In hips free of definite RHOA, our AUCs for the incident RHOA models showed good predictive performance in similar populations. However, the added predictive value of the morphology measurements was small, and model performance was heterogeneous in leave-one-cohort-out cross-validation.

Initial results of this work have been previously presented at the 2024 Osteoarthritis Research Society International (OARSI) World Congress on Osteoarthritis; Vienna, Austria; April 14–21, 2024 (https://doi.org/10.1016/j.joca.2024.02.105), Combined 61st NOF Congress and NOV Jaarcongres 2024; Rotterdam, June 12–14, 2024; and the 18th International Workshop on Osteoarthritis Imaging; Marrakesh, Morocco; June 25–28, 2024 (https://doi.org/10.1016/j.ostima.2024.100212).

The World COACH consortium has been funded through grants by the Dutch Arthritis Society (grant 21-1-205, LLP34), the Dutch Research Council (now; Veni: 09150161910071), and Erasmus MC University Medical Center Rotterdam. This publication is part of the Flagship Healthy Joints, which is (partly) financed by Convergence Health and Technology. Dr Lindner's work was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant 223267/Z/21/Z). This research was funded in whole, or in part, by the Wellcome Trust (grant 223267/Z/21/Z).

[1]Myrthe A. van den Berg, MSc, Fleur Boel, MSc, Michiel M. A. van Buuren, MD, Noortje S. Riedstra, MD, PhD, Jinchi Tang, MD, PhD, Sita M. A. Bierma-Zeinstra, PhD, Cindy G. Boer, PhD, Joyce B. J. van Meurs, PhD, Edwin H. Oei, PhD, Jos Runhaar, PhD, Rintje Agricola, MD, PhD: Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; [2]Harbeer Ahedi, PhD: University of Tasmania, Menzies Institute for Medical Research, Hobart, Tasmania, and Sydney Orthopaedic Research Institute, Sydney, New South Wales, Australia; [3]Nigel K. Arden, MD: University of Oxford, Botnar Research Centre, Oxford, United Kingdom; [4]Flavia M. Cicuttini, PhD: Monash University, Melbourne, Victoria, Australia; [5]Timothy F. Cootes, PhD, Claudia Lindner, PhD: The University of Manchester, Manchester, United Kingdom; [6]David T. Felson, MD: Boston University School of Medicine, Boston, Massachusetts; [7]Willem Paul Gielis, MD, PhD: University Medical Center Utrecht, Utrecht, The Netherlands; [8]Joshua Heerey PhD, Andrea Mosler, PhD: La Trobe University, Melbourne, Victoria, Australia; [9]Graeme Jones, MD: University of Tasmania,

SIGNIFICANCE & INNOVATIONS

- Our work on hip morphology–based osteoarthritis risk prediction modeling uses an individual participant data meta-analysis to overcome the limitations of small sample sizes in single-cohort designs in current epidemiologic research focused on the role of single-type (eg, cam or pincer) hip morphology on radiographic hip osteoarthritis risk.
- Combining data from 35,984 hips from all available prospective cohort studies with consecutive hip imaging data worldwide (via World COACH), our work uses advanced statistical models to evaluate the importance of continuous hip morphology measures on radiographic hip osteoarthritis (RHOA) risk.
- Although hip morphology was expected to enhance identification of individuals at risk for RHOA, based on its previously established association with RHOA development in isolation, our findings showed that these measures did not improve risk prediction beyond baseline demographics and RHOA grades, highlighting both the complexity of RHOA incidence and challenges in modeling across heterogeneous cohorts.
- Our research demonstrates the importance of large-scale meta-analyses for risk modeling and highlights the significance of properly validating risk models in various populations for future clinical application.

INTRODUCTION

Despite the growing burden of hip osteoarthritis (HOA) on society and health care systems, prevention and treatment approaches are still only slowly emerging.[1] Early identification of HOA is important for improving clinical decision-making and advancing our understanding of disease development and potential prevention and treatment options.[2] By gaining insights into risk factors and thereby identifying individuals with a high risk of developing HOA, more specific management strategies could be designed for specific subgroups.

Previous research has established that hip morphology plays a key role in the development of radiographic HOA (RHOA).[3,4] Hip morphology–related conditions that have been shown to increase the risk of developing RHOA include acetabular dysplasia, pincer morphology, and cam morphology. In research settings, these hip morphologies are assessed using cutoff values of continuous morphologic measurements.[5] However, it is currently recommended to avoid these cutoff values and use the continuous measurements when possible.[6,7]

Current RHOA risk prediction models based on morphology are limited in the level of flexibility and generalizability to unseen data. This problem is caused by the relatively small size of the available datasets and varying study designs.[8] Therefore, most studies have focused on the risk posed by one type of hip morphology determined by conventional statistical methods.

A larger dataset would allow the investigation of linear and nonlinear interactions between RHOA risk and multiple continuous morphology measurements with flexible statistical and machine learning models. A promising solution to increase the sample size is to combine individual participant data (IPD) of various studies while taking differences between studies into account. These differences could also provide insight into the generalizability of the developed prediction models during internal and external model validation. This approach is expected to increase our understanding of the effect of identified hip morphologies as risk factors on the development of RHOA and could result in more robust prediction models.[9,10] Additionally, this approach could show if automatic and interpretable assessment of hip morphology on radiographs could enable scalable risk identification and facilitate implementation in everyday clinical settings.

This study aimed to develop a hip morphology–based RHOA risk prediction model by investigating whether including hip morphology measurements could improve the ability of a model to distinguish people with and without RHOA risk within four to eight years. With the use of a large multicohort dataset, we tested the generalizability of our results in both similar and unseen populations.

MATERIAL AND METHODS

Study design and participants. This study was designed and reported according to the transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster) checklist.[11] Study participants were selected from the Worldwide Collaboration on OsteoArthritis prediCtion for the Hip (World COACH) consortium. This consortium combines data from prospective cohort studies worldwide with sequential pelvic or hip imaging available. Details of this consortium are published elsewhere.[12] The current study first

Menzies Institute for Medical Research, Hobart, Tasmania, Australia; [10]Stefan Kluzek, PhD: University of Oxford, Botnar Research Centre, Oxford, and University of Nottingham, Nottingham, United Kingdom; [11]Nancy E. Lane, MD: University of California at Davis, Davis; [12]Amanda E. Nelson, MD: University of North Carolina at Chapel Hill, Chapel Hill; [13]Michael C. Nevitt, PhD: University of California, San Francisco; [14]Harrie Weinans, PhD: University Medical Center Utrecht, Utrecht, and Delft University of Technology, Delft, The Netherlands; [15]Jesse H. Krijthe, PhD: Delft University of Technology, Delft, The Netherlands.

Author disclosures and a graphical abstract are available online at https://onlinelibrary.wiley.com/doi/10.1002/acr.25629.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25629).

Address correspondence via email to Rintje Agricola, MD, PhD, at r.agricola@erasmusmc.nl.

Submitted for publication January 8, 2025; accepted in revised form July 31, 2025.

considered cohorts with standardized anteroposterior (AP) pelvic, long-limb, and/or hip radiographs taken and graded for RHOA at baseline and four to eight years of follow-up. This resulted in the initial inclusion of 38,594 participants from nine cohorts (ie, Cohort Hip and Cohort Knee [CHECK], the Chingford Study, The Johnston County Project [JoCoOA], Multicenter Osteoarthritis Study [MOST], OsteoArthritis Initiative [OAI], Rotterdam Study [RS-I, RS-II, RS-III], and the Study of Osteoporotic Fractures [SOF]).

Additional inclusion criteria were applied to IPD at the hip level. Baseline AP pelvic radiographs were required to have sufficient quality to determine pelvic landmark points for automatic hip morphology assessment. Hips only shown in AP hip (rather than pelvic) radiographs were excluded because a horizontal reference line, needed for adjustment for pelvic tilt, requires both hips in view. Additionally, hips must have been graded for RHOA grade at baseline and follow-up visit by the original cohort. All hips with the outcomes (ie, definite RHOA or a total hip replacement [THR] at baseline) were also excluded. In total, 35,984 hips (47%, 18,854 participants) with complete data were included in this study.

Prediction model definition. *Outcome definition and considered risk factors.* RHOA grades were obtained from the original cohort data, all of which used either the Kellgren and Lawrence (KL) or modified Croft grading systems.[13,14] These scores were harmonized to a three-grade system: free of RHOA (KL or Croft = 0), doubtful RHOA (KL or Croft = 1), and definite RHOA (KL or Croft ≥2 or THR). Incident RHOA was defined as hips with no definite RHOA at baseline that developed definite RHOA or had a THR within four to eight years. Our baseline population, therefore, included hips with no signs of RHOA (RHOA grade = 0) or doubtful RHOA (RHOA grade = 1).

To assess hip morphology as a risk factor, the lateral center edge angle (LCEA) and alpha angle (AA) were determined uniformly on the baseline radiographs of all included cohorts. Using the BoneFinder software (www.bone-finder.com; The University of Manchester, UK), the outline of the proximal femur and acetabulum was automatically annotated with a custom-made search model.[15,16] Subsequently, this point-set facilitated the automated LCEA and AA measurements through a Python script. Details regarding the reliability and validity of this method have been published previously.[17] The intermethod intraclass correlation coefficients (ICCs) with 95% confidence interval (CI) between manual and automated measurements were 0.89 (95% CI 0.78–0.94) for the LCEA and 0.46 (95% CI 0.12–0.70) for the AA, similar to the found interobserver ICCs.[16]

The LCEA measurement is depicted in Figure 1A. This measurement quantifies the coverage of the femoral head by the acetabulum and is corrected for pelvic tilt. Both under coverage by the acetabulum (smaller LCEA) and over coverage by the acetabulum (larger LCEA) could be potential risk factors for incident RHOA.[3,18–25] Therefore, we expected a nonlinear relationship between the continuous LCEA measurement and incident RHOA risk.

The AA measurement is depicted in Figure 1B. This measurement describes the degree of sphericity of the lateral femoral head-neck junction and is used to detect the presence and quantify the degree of cam morphology. A linear relationship between the continuous AA and RHOA is expected, such that the higher the AA (lower sphericity), the higher the risk of incident RHOA.[3,22,23,26,27]

Four different models were defined. Model 1 included demographic information, hip side, and follow-up time. The demographic information included baseline age, body mass index, and sex assigned at birth, as they are previously proposed risk factors for incident RHOA.[28] Hip side (left or right hip) and follow-up time in years were included to account for our hip-level modeling approach and adjusted for the different follow-up times of the different cohorts, potentially imposing differences in RHOA risk. In model 2, baseline RHOA grade was added to model 1 to adjust for the expected risk difference for hips with no (RHOA grade 0) and doubtful (RHOA grade 1) RHOA. This risk difference is expected, as hips having an RHOA grade of 1 already show early radiographic changes that can lead to RHOA. After subsequently adding the LCEA and AA measurements in model 3, we considered RHOA grading approach (KL or Croft), cohort location (United States or Europe), and cohort type (open or closed population) as cohort-describing variables in model 4.

Risk prediction modeling approaches. The dataset used is a multilevel dataset, meaning that the data of participants is clustered within cohorts with different cohort-specific characteristics. In this IPD meta-analysis, we considered the correlation among hips from the same cohort in our modeling approaches. One of the conventional approaches to handling multilevel data in risk prediction is generalized linear mixed-effects models (GLMM) for a binary outcome.[29] For the present study, different GLMMs were built with a binomial distribution and a logit link function. We defined a nested random intercept to account for the correlation of measurements in hips within a participant within a cohort. Ultimately, the estimated marginal model intercept was shifted with a combination of a participant-specific and cohort-specific offset value for the participants the model was built on. When new data from the same cohort were used in model testing, the estimated intercept was adjusted with this cohort-specific adjustment. For participants from new cohorts, the estimated marginal model intercept was used. We considered all risk factors to have a common effect across cohorts and to have no interactions with each other.

As an alternative approach, to step away from the assumptions made in a GLMM model structure, we also built random forest (RF) models that considered cohort variability by including the cohort label as a risk factor. An RF model combines the output of multiple decision trees, making it a nonlinear modeling approach that considers all possible interactions between risk factors.

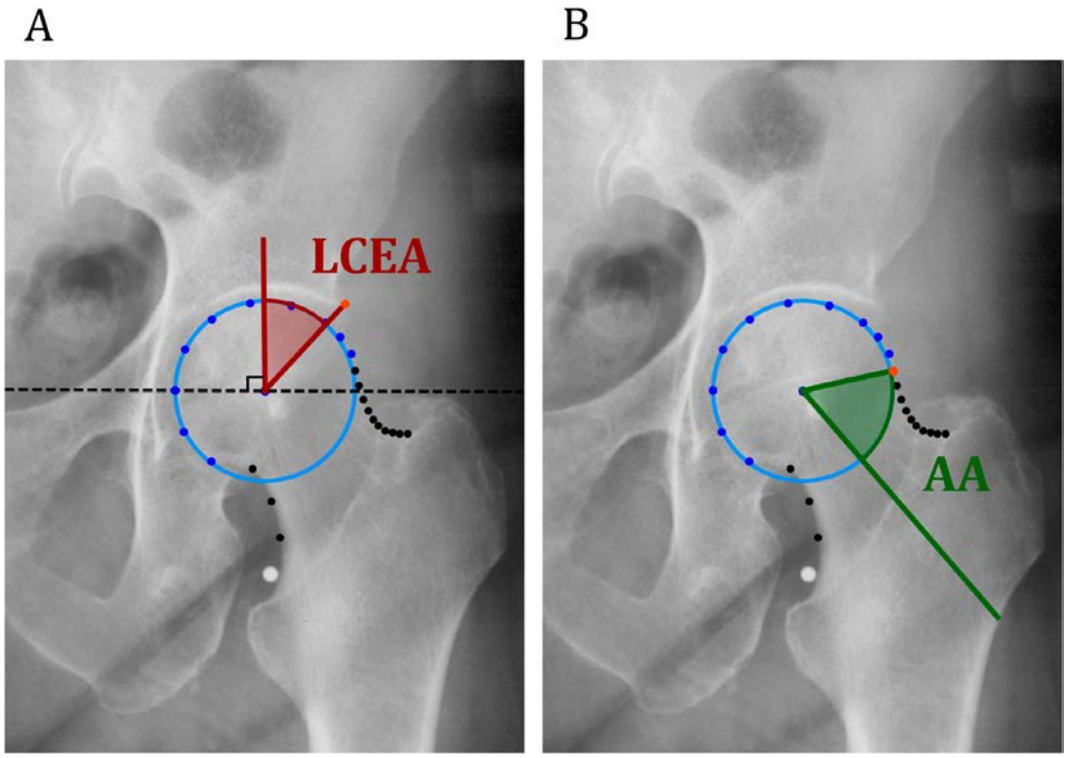

Figure 1. Anteroposterior pelvic radiographs focused on the left hip showing the automatically measured morphology measurements. These methods rely on automated landmark points. To determine the center of the femoral head, a best-fitting circle (light blue) is outlined around the femoral head based on a subset of femoral head landmark points (dark blue). (A) LCEA (in red): The angle is calculated between a line drawn vertically through the center of the femoral head, perpendicular to the horizontal reference line (black, dashed), and a line originating from the center of the femoral head to the most lateral part of the acetabulum (orange landmark). (B) AA (in green): The angle is calculated between a line through the femoral head center and the point where the femoral head or femoral neck is first outside of the best-fitting circle (orange landmark) and the femoral neck axis. AA, alpha angle; LCEA, lateral center edge angle.

It can also model a nonlinear relationship between the risk factors and incident RHOA.[30] For individuals from new cohorts, these variables indicating cohort participation would all be set to 0. After initial experiments, hyperparameters on the RF models were set to a maximum depth of 10, number of estimators of 200, and a minimum of eight samples per leaf.

Statistical analysis. Differences among baseline characteristics of the included and excluded hips and between cohorts were investigated with descriptive statistics. The continuous variables were standardized based on the mean and SD determined on the full dataset before fitting the models. Outcomes were reported based on the original unit scale. Two primary analyses were performed to (1) investigate the associations between the considered risk factors and incident RHOA and to determine the discriminative performance of multiple GLMM definitions; and (2) validate the model performance of both GLMMs and RF models on unseen datasets.

(1) Associations between risk factors and incident RHOA and determining the discriminative performance of GLMM definitions. We first compared the adjusted odds ratios (aOR) with 95% CI of the risk factors in four different GLMM specifications. All GLMMs were built with the lme4 package in R version 4.1.1.[31] The assumed nonlinear effect of the LCEA measurement was modeled with natural cubic splines with two degrees of freedom. Model performance was assessed using stratified five-fold cross-validation, in which the relative contribution to the dataset of a cohort and the percentage of hips at risk were kept constant. Therefore, information for cohort-specific adjustment was available. The discriminative performance of models with or without a cohort-specific intercept was compared. The area under the receiver operating characteristic curve (AUC) was reported as the mean and SD of the measurement on the five resulting test sets. The AUC classifies the predictive performance as fail ($0.5 \leq AUC < 0.6$), poor ($0.6 \leq AUC < 0.7$), fair ($0.7 \leq AUC < 0.8$), good ($0.8 \leq AUC < 0.9$), or excellent ($0.9 \leq AUC$).[32]

(2) Validation of GLMM and RF models on unseen datasets. To assess the discriminative and calibration performance of both the GLMMs and RF models including the hip morphology measurements to unseen populations, leave-one-cohort-out cross-validation was used. Each time, a model was built on a dataset of eight cohorts and evaluated on the left-out cohort. The GLMMs

were built with the lme4 package in R version 4.1.1.[31] The RF models were built with the RandomForestClassifier function in Scikit-learn version 1.2.0, within Python version 3.9.[33] The discriminative performance was evaluated by the mean AUC value and 95% CI in the nine left-out cohorts. Additionally, the calibration statistics of the GLMMs and RF models were compared in each left-out cohort.

Data sharing statement. Data are available upon reasonable request. Data may be obtained from a third party and are not publicly available. We encourage the use of data by third parties, although this is subject to approval by the steering committees of the World COACH consortium and the participating cohorts, as well as to legal boundaries regarding data ownership. A standardized data request form is available for which will be reviewed uniformly in order to consistently handle World COACH data requests.

Patient and public involvement. The World COACH consortium encourages the public to share ideas, stay up to date on our findings, and ask questions through the website (www.worldcoachconsortium.com). Additionally, we help organize annual conferences for patients with OA in the Netherlands (Artrose Gezond, https://artrosegezond.nl/). Patients and public were involved in the design, conduct, reporting, and/or dissemination plans of this research. OA representatives were also involved in setting up the consortium goals and research questions.

Ethical approval. This study involves human participants but was excepted from ethical approval (Erasmus MC Medical Ethics Review Committee) as it uses previously collected observational data for which the participants had originally given informed consent, and all cohort studies included in this consortium already had ethics approval from their respective committees. Participants gave informed consent to participate in the study before taking part.

RESULTS

Participants. A detailed overview of the flow of inclusion and exclusion of hips while pooling the data of the nine cohorts to the final included set is shown in Figure 2. The largest proportion (62%) of hips were excluded because these hips were not graded for RHOA at baseline by the cohort or had no radiograph and/or RHOA grade available at follow-up. Descriptive statistics of the pooled included and excluded dataset, stratified by cohort, are given in Table 1. The excluded population was older than the included population (67.5 ± 10.6 years compared to 63.5 ± 8.5 years). The other descriptive characteristics were similar.

In our included set of 35,984 hips free of definite RHOA at baseline, 1,690 hips developed incident RHOA within four to eight years. Whereas the pooled prevalence of incident RHOA was 4.7%, the prevalence among the cohorts ranged from 1.5% for OAI to 20.2% for CHECK. Additionally, the prevalence of doubtful RHOA ranged from 9% in RS-III to 76% in JoCoOA at baseline. Other differences between the cohort populations at baseline were their mean baseline age, ranging from 54 to 71 years, and sex distribution, with Chingford and SOF being female-only cohorts.

Associations of risk factors and incident RHOA and discriminative performance of GLMM definitions. The aORs and discriminative performance in terms of AUC for our four GLMM definitions are shown in Table 2. Summary statistics on the data used for each train and test dataset combination and model specifications can be found in the Supplementary Material. For the LCEA, the two values of the aOR represent the exponent of the first and second components of the natural cubic spline transformation, respectively. To interpret the effect of 1° of change in the LCEA value, the predictor effect plot for GLMM3 is shown in Figure 3. An aOR of 1.018 for the AA can be interpreted as that, for every 1° increase in the AA value, the odds of incident RHOA increase by 1.8%. Similarly, having an LCEA of 20° compared to 21°, the odds of incident RHOA decrease by 8% (aOR = 0.924).

When using the cohort-specific intercept, adding the baseline RHOA grade (ie, GLMM1 vs GLMM2) improved the discriminative performance of the model from a mean AUC of 0.72 to 0.80 (Table 2), whereas the morphology measurements (GLMM3) did not change this performance. Overall, the discriminative performance was ∼0.1 in AUC value higher when using a cohort-specific intercept to predict the risk of incident RHOA in GLMM1, GLMM2, and GLMM3. When adding the cohort descriptive variables as fixed-effects in the model, the AUC was reduced to 0.76 using the cohort-specific intercept but increased to 0.75 with the marginal intercept, compared to the other model definitions. In GLMM4, the cohort-specific adjustments to the intercept were found to be nonexistent. However, the cohort-specific variables did not show a statistically significant association with incident RHOA. Adding these variables resulted in different aOR values and larger CIs for the other considered risk factors. Specifically, the aOR for having doubtful RHOA at baseline was 17.8. Therefore, the results for GLMM4 should be interpreted with caution.

GLMM and RF model validation on unseen datasets. Table 3 shows the results of the leave-one-cohort-out validation for both the GLMMs and RF models. These models included the baseline patient characteristics, RHOA grade, and morphology measurements (as defined in GLMM3). The summary statistics on the data used for each train and test dataset combination, model specifications, calibration curves, and receiver operating characteristic curves can be found in the Supplementary Material.

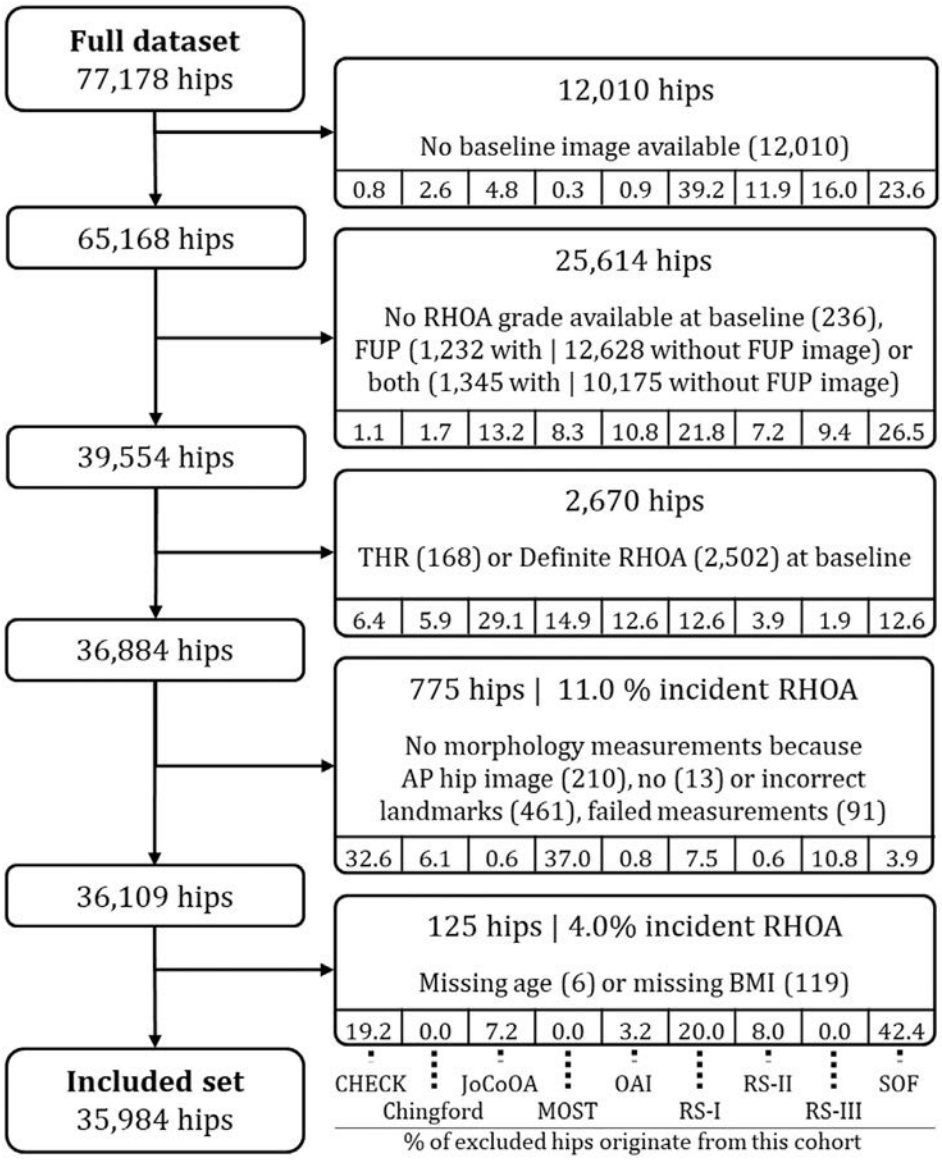

Figure 2. Flow of hips from the fully pooled dataset of the nine cohorts to the final included set. AP, anteroposterior; BMI, body mass index; CHECK, Cohort Hip and Cohort Knee; Chingford, Chingford Study; FUP, follow-up visit; JoCoOA, Johnston County Project; MOST, Multicenter Osteoarthritis Study; OAI, OsteoArthritis Initiative; RHOA, Radiographic hip osteoarthritis; RS-I, Rotterdam Study-I; RS-III, Rotterdam Study-III; SOF, Study of Osteoporotic Fractures; Study-II RS-II, Rotterdam; THR, total hip replacement.

The discriminative performance (AUC) ranged from 0.57 to 0.88 and 0.56 to 0.87 for the GLMMs and RF models, respectively. Validation on the OAI and RS-II cohorts had the highest AUC values, which were the cohorts with the lowest prevalence of incident RHOA. Validation on Chingford and JoCoOA showed the worst performance. The mean predicted values were mostly higher for the RF model, except for RS-I and SOF. When analyzing the model calibration, we observed that models overestimated the risk for the same two cohorts. For CHECK, JoCoOA, Chingford, and RS-III, both models underestimated the risk. For MOST, OAI, and RS-II, the GLMM underestimated and the RF model overestimated the risk. The RF models seemed better calibrated than the GLMM.

DISCUSSION

We thoroughly investigated the effect of two continuous hip morphology measurements on incident RHOA within four to eight years in a large international IPD analysis of nine cohorts and nearly 36,000 hips. During stratified cross-validation, similar discriminative performance was found for models with and without the LCEA and AA measurements (mean AUC 0.80 ± 0.01).

Table 1. Descriptive statistics of the included cohorts, the pooled included set, and pooled excluded set*

	CHECK	Chingford	JoCoOA	MOST	OAI	RS-I	RS-II	RS-III	SOF	Included set	Excluded set[a]
Participants											
Excluded participants, n	361	408	2,212	1,310	1,494	5,260	1,671	2,207	4,812		19,735
Included participants, n	641	595	1,772	1,716	3,302	2,860	1,341	1,732	4,895	18,854	
Sample size											
Hips, n	1,189	1,060	3,209	3,197	6,372	5,542	2,624	3,408	9,383	35,984	41,194
Right hips, %	49.9	47.0	51.6	50.4	49.9	49.7	49.6	50.1	49.6	49.9	50.1
Baseline factors											
Age at baseline, mean ± SD, y	55.8 ± 5.22	53.5 ± 5.71	60.1 ± 8.68	61.7 ± 7.72	60.7 ± 9.03	65.3 ± 6.52	63.2 ± 6.56	56.3 ± 5.01	70.6 ± 4.50	63.4 ± 8.47	67.5 ± 10.58
Male hips, n (%)	210 (17.7)	0 (0.0)	1,232 (38.4)	1,096 (34.3)	2,704 (42.4)	2,412 (43.5)	1,159 (44.2)	1,507 (44.2)	0 (0.0)	10,320 (28.7)	11,750 (28.5)
BMI at baseline, mean ± SD, kg/m^2	26.1 ± 4.00	25.4 ± 3.99	29.5 ± 5.88	29.5 ± 4.85	28.2 ± 4.57	26.3 ± 3.47	27.1 ± 3.90	27.5 ± 4.13	26.5 ± 4.36	27.4 ± 4.57	27.7 ± 5.29
Hips with RHOA grade 1, n (%)	293 (24.6)	71 (6.7)	2,445 (76.2)	1,193 (37.3)	809 (12.7)	2.203 (39.8)	372 (14.2)	322 (9.4)	4,583 (48.8)	12,291 (34.2)	5,784 (27.7)
AA, mean ± SD, degrees	46.4 ± 9.92	51.3 ± 12.63	44.7 ± 8.64	47.1 ± 9.64	48.1 ± 11.45	46.4 ± 10.13	45.5 ± 9.52	50.8 ± 12.84	45.1 ± 9.57	46.7 ± 10.57	47.7 ± 11.75
LCEA, mean ± SD, degrees	34.7 ± 5.59	36.1 ± 6.04	36.3 ± 5.86	33.4 ± 6.05	35.3 ± 5.64	36.6 ± 5.81	35.5 ± 5.68	34.7 ± 5.62	38.2 ± 5.89	36.1 ± 5.99	36.7 ± 6.35
Cohort-level covariates											
Follow-up time, y	8	7	6	5	4	8	4	6	8	4–8	4–8
RHOA grading system	KL	KL	KL	KL	Croft	KL	KL	KL	Croft	KL, Croft	KL, Croft
Location of cohort	Europe	Europe	United States	United States	United States	Europe	Europe	Europe	United States	Europe, United States	Europe, United States
Included population type	Closed	Open	Open	Closed	Closed	Open	Open	Open	Open	Open, closed	Open, closed
Outcome of interest											
Hips with incident RHOA, n (%)	240 (20.2)	110 (10.4)	386 (12.0)	133 (4.2)	95 (1.5)	173 (3.1)	44 (1.7)	146 (4.3)	363 (3.9)	1,690 (4.7)	94 (10.4)

* AA, alpha angle; BMI, body mass index; CHECK, Cohort Hip and Cohort Knee; Chingford, Chingford Study; JoCoOA, Johnston County; KL, Kellgren and Lawrence grade; LCEA, lateral center edge angle; MOST, Multicenter Osteoarthritis Study; OAI, OsteoArthritis Initiative; RHOA, radiographic hip osteoarthritis; RS-I, Rotterdam Study-I; RS-II, Rotterdam Study-II; RS-III, Rotterdam Study-III; SOF, Study of Osteoporotic Fractures.

[a] Sample sizes vary for the descriptive statistics of the excluded set as they are based on only the available data for this variable.

Table 2. Risk factor contribution in terms of aORs of four different multivariate GLMM definitions on the pooled included dataset and their discriminative performance based on stratified cross-validation*

	GLMM1	GLMM2	GLMM3	GLMM4
Age, in years	**1.051 (1.041–1.062)**	**1.043 (1.033–1.053)**	**1.043 (1.033–1.054)**	1.011 (0.947–1.079)
Sex at birth				
Female [reference]				
Male	1.118 (0.954–1.310)	0.979 (0.830–1.155)	**0.828 (0.695–0.987)**	0.389 (0.113–1.336)
BMI, in kg/m^2	1.012 (0.998–1.026)	1.012 (0.998–1.027)	1.009 (0.994–1.024)	1.023 (0.920–1.138)
Hip side				
Left [reference]				
Right	1.101 (0.987–1.229)	**1.149 (1.025–1.290)**	**1.158 (1.031–1.301)**	**1.476 (1.206–1.806)**
Follow-up time, in years	**1.511 (1.124–2.032)**	1.410 (0.928–2.144)	1.409 (0.942–2.106)	1.212 (0.764–1.923)
RHOA grade				
Free of RHOA [reference]				
Doubtful RHOA		**9.000 (7.602–10.655)**	**8.751 (7.378–10.38)**	**17.820 (11.768–26.985)**
AA, in degrees			**1.018 (1.012–1.024)**	1.018 (0.999–1.038)
LCEA,[a] in degrees				
First spline component			**0.062 (0.019–0.197)**	0.009 (0.000–0.390)
Second spline component			1.771 (0.856–3.662)	4.931 (0.528–46.029)
Scoring approach				
KL [reference]				
Croft				0.353 (0.074–1.677)
Cohort type				
Open [reference]				
Closed				2.593 (0.775–8.672)
Location				
Europe [reference]				0.483 (0.115–2.029)
United States				
AUC				
With cohort-specific intercept	0.72 ± 0.01	0.80 ± 0.01	0.80 ± 0.01	0.76 ± 0.02
With marginal intercept	0.56 ± 0.01	0.70 ± 0.01	0.71 ± 0.01	0.75 ± 0.01

* Values are in aOR (95% confidence interval) or mean ± SD. Bold numbers indicate a significant association ($P < 0.05$). AA, alpha angle; aOR, adjusted odds ratio; AUC, area under the receiver operating characteristic curve; BMI, body mass index; GLMM, generalized linear mixed model; KL, Kellgren and Lawrence; LCEA, lateral center edge angle; RHOA, radiographic hip osteoarthritis.
[a] Variable is modeled with natural cubic splines (*df* = 2), displayed aORs are the exponent of the beta values of the two spline components.

When comparing GLMM and RF modeling strategies and their generalizability to unseen populations, similar performances were found with AUC values ranging between 0.56 and 0.88 for each left-out cohort. This implies that using a more flexible model definition that includes possible risk factor interactions or nonlinear relationships would not improve the models. The RF models were better calibrated, which implies that the estimated probabilities are more consistent with the observed RHOA risk distribution within the population.

Previous studies examining the role of hip shape in RHOA risk primarily converted the measurements into a binary risk factor using specific thresholds. Considering hip morphology measurements as a continuous risk factor allows for a more data-driven estimation of the effect. Despite recognizing several hip morphologies as RHOA risk factors, inconsistencies in the literature regarding the thresholds used complicate the interpretation of reported associations.[34] In agreement with previous reports, we found increased odds of incident RHOA when the morphologies were present.[21–23,35]

In contrast, we found comparable AUC values when evaluating the discriminative performance of the models with and without hip morphology measurements (GLMM3 vs GLMM2). This indicates that inclusion of the AA and LCEA did not improve the ability of the model to distinguish between hips that developed or did not develop incident RHOA. Possible explanations are that these geometric measurements alone are not descriptive enough to capture the structural differences between individuals at and not at risk for RHOA development. The low to moderate interrater and intermethod ICCs for the AA might also play a role, potentially reducing the sensitivity of the measurement to identify individuals at risk. Alternatively, abnormal hip morphology may be a true risk factor for early-onset hip OA, but the high-risk participants are already excluded from this analytic population.[23]

The largest increase in AUC values was observed when adding the baseline RHOA status to our model. This indicates that doubtful or mild radiographic changes (ie, RHOA grade 1) identified hips at risk for incident RHOA within four to eight years. Our findings are comparable to those in Saberi Hosnijeh et al, in which an increase of 0.11 in AUC on the training set (RS-I) was found when the imaging variables baseline KL score, thumb OA, hip dysplasia, and cam morphology were added to their basic model; a similar increase in AUC values was found in external validation on CHECK and RS-II data.[36] Based on these findings, in combination with our results, baseline RHOA grade was expectedly the largest contributor to this increase in AUC. Additionally, where Saberi Hosnijeh et al showed an aOR of 4.43 (3.14–5.77) for

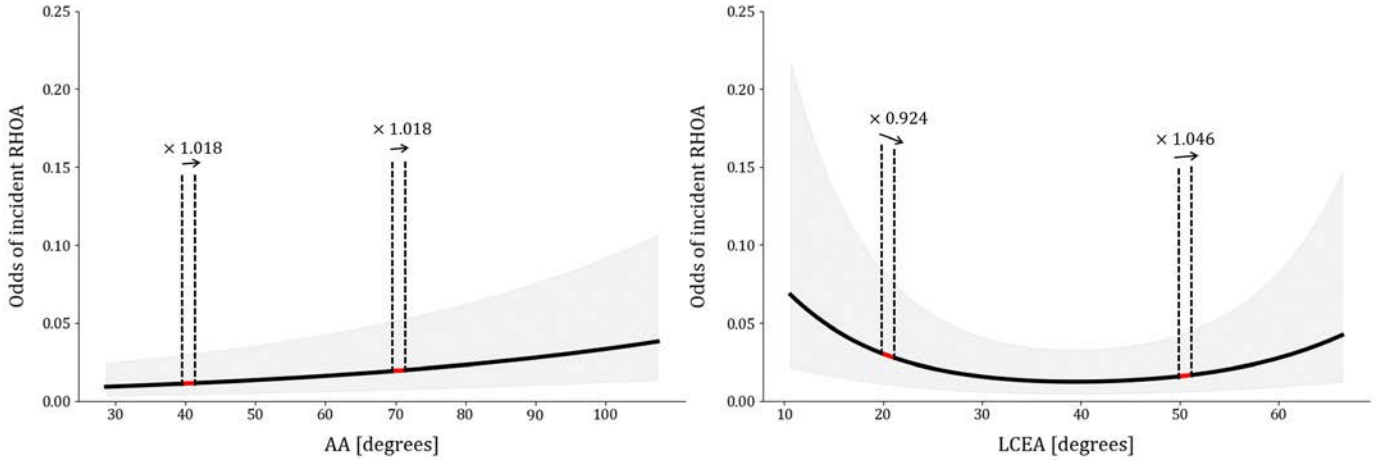

Figure 3. Predictor effect plots of the morphology measurements included in the multivariable risk-prediction model GLMM3. Odds are given for a female participant after eight years of follow-up, with the mean values of the continuous factors and the reference values for the categorical factors. The annotated multiplication factors describe the change in odds for an increase of 1° at that location. AA, alpha angle; LCEA, lateral center edge angle; GLMM, generalized linear mixed model; RHOA, radiographic hip osteoarthritis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25629/abstract.

having KL grade 1, we found an aOR of 8.75 (7.38–10.38) for having either KL or Croft grade 1 in GLMM3. Although various opinions exist on considering no versus doubtful RHOA within models for incident RHOA to better understand the specific role of risk factors, including baseline RHOA status in prediction models better identifies hips at risk.

Additionally, differences in performance were found when a cohort-specific intercept instead of a marginal intercept was used. This implies that the currently considered variables do not explain the heterogeneity in incident RHOA prevalence between the cohorts. When cohort-level covariates were added to account for variability in inclusion criteria and cohort setup, the marginal and cohort-specific intercepts were identical. However, because the number of cohorts is always going to be limited, it is inherently difficult to estimate the cohort-level parameters accurately. We hypothesize that the heterogeneity in incident RHOA prevalence is mainly the result of the differences in cohort setup (ie, follow-up time and specific inclusion of people at risk of OA), the semiobjective nature of the RHOA scoring systems, and the different scoring approaches used by the cohorts.[37] Future work in the World COACH consortium aims to include standardized RHOA scoring across cohorts to ensure that observed heterogeneity reflects actual population differences rather than inconsistencies in scoring methods.

The heterogeneity between cohorts contributed to heterogeneous model performance for leave-one-cohort-out validation. Overall, we observed that when the prevalence in the training dataset was higher than the prevalence in the test dataset, discriminative performance improved. As noted by Riley et al, even when the found predictor effects are consistent, variations in characteristics or prevalence across different settings or populations can lead to genuine differences in model calibration due to

Table 3. Discriminative performance and mean predicted probabilities of the GLMM and RF models based on leave-one-cohort-out cross-validation*

	Risk of incident RHOA		GLMM		RF	
Cohort	OA/No-OA	Prevalence, %	Mean predicted values, %	AUC (95% CI)	Mean predicted values, %	AUC (95% CI)
CHECK	240/949	20.2	0.01	0.69 (0.65–0.73)	3.24	0.68 (0.64–0.72)
Chingford	110/950	10.4	0.91	0.57 (0.51–0.63)	2.26	0.57 (0.51–0.63)
JoCoOA	386/2,823	12.0	4.53	0.57 (0.54–0.60)	7.02	0.56 (0.53–0.59)
MOST	133/3,064	4.2	2.25	0.80 (0.77–0.83)	5.33	0.80 (0.76–0.83)
OAI	95/6,277	1.5	0.88	0.86 (0.82–0.90)	3.60	0.82 (0.77–0.87)
RS-I	173/5,369	3.1	9.72	0.78 (0.75–0.80)	5.99	0.76 (0.73–0.79)
RS-II	44/2,580	1.7	0.97	0.88 (0.83–0.94)	3.01	0.87 (0.80–0.93)
RS-III	146/3,262	4.3	1.04	0.75 (0.70–0.80)	2.44	0.74 (0.69–0.79)
SOF	363/9,020	3.9	14.61	0.69 (0.66–0.71)	7.16	0.69 (0.67–0.72)

* AUC, area under the receiver operating characteristic curve; CHECK, Cohort Hip and Cohort Knee; Chingford, Chingford Study; CI, confidence interval; GLMM, generalized linear mixed model; JoCoOA, Johnston County Project; MOST, Multicenter Osteoarthritis Study; OAI, OsteoArthritis Initiative; RF, random forest; RHOA, radiographic hip osteoarthritis; RS-I, Rotterdam Study-I; RS-II, Rotterdam Study-II; RS-III, RS-III, Rotterdam Study-III; SOF, Study of Osteoporotic Fractures.

incorrect estimation of the intercept.[38] In general, an underestimation of incident RHOA risk was found, except for the two older cohorts RS-I and SOF. This could indicate that the considered risk factors have a different effect on RHOA risk for this older population than in the younger training population. More efforts in modeling and understanding the underlying factors for incident RHOA prevalence in these different cohorts could therefore improve RHOA risk estimates. Despite our efforts to combine data from multiple cohorts via IPD to consider a more versatile population to better estimate the intercept and risk factors in new populations, we observed spectrum bias.[39] In addition to creating a larger dataset, the target population should be considered when selecting cohorts to combine, and more risk factors should be added to explain any variability between the cohorts.

The major strength of this study is that it is the first to use data from all available prospective cohort studies on HOA to create risk prediction models for incident RHOA within four to eight years. By combining these IPD, we can include a large and heterogeneous study population. This covers the general population of interest better than individual single-cohort studies. By uniformly measuring our baseline factors, automatically determining our morphology measurements, and taking the cohort differences into account in our modeling approaches, we could provide a more precise estimation of the performance of our prediction models.[40] Through evaluation of models via both internal and external validation, we assessed the generalizability of our predictions, which, although highly encouraged, can be challenging for clinical risk prediction models.

The limitations of the study are mainly caused by the differences in setup of the nine included cohorts, such as varying imaging protocols, inclusion criteria, and follow-up times. Firstly, our defined follow-up time of four to eight years might have been too short and too broad to determine reliable RHOA risk estimates due to the low prevalence of incident RHOA, and this variability in follow-up duration could reduce the apparent strength of associations. When considering an extended follow-up for cohorts with this data available, we see that the prevalence of incident RHOA can double when looking, for example, at 11 years of follow-up for RS-I and RS-II. Secondly, by only defining hip morphology on AP radiographs, we might underestimate its presence and severity, which is better captured in three dimensions. Nevertheless, these automated measurements on two-dimensional imaging allow for easier screening of large populations. Additionally, as the inclusion criteria were applied at the hip level, one potential unadjusted and unmeasured confounder was the increased risk of incident RHOA in individuals with contralateral RHOA. However, this was not the focus of the study. Moreover, including THR in our definition of incident RHOA might have led to some misclassification, as individuals could have undergone THR for other reasons than hip OA. Nonetheless, this group of people accounts for less than 5% of all incident RHOA cases in our study. Lastly, our outcome definition of incident RHOA was dependent on semiobjective scoring systems with high intra- and interrater variability,[41] and Fan et al have shown that prevalence estimates with the KL and Croft grading systems were statistically different.[1] Although the defined risk of interest aligns with the current clinical practice, it does limit the interpretability of the results.

Based on the current study, using more automated RHOA grading systems should be encouraged in the future to reduce variability in scoring for multicohort analyses. In addition, automated and uniform RHOA grading methods could reduce the time needed to reliably grade radiographs already collected in epidemiologic research and could therefore increase the sample size. To aid clinical decision-making and treatment options, multimodal prediction modeling with risk factors from different domains, such as genetics and clinical examination, is recommended.[42] Although the set of considered risk factors in this study already provided valuable insights into incident RHOA, creating artificial intelligence models that can analyze radiographs and consider many more than only demographic and hip morphology–based risk factors could improve our understanding of HOA development. This might also reduce the present heterogeneity in model performance when applied to new populations.

In conclusion, the hip morphology–based risk prediction models built with multicohort data had good predictive performance for incident RHOA within four to eight years for hips free of definite RHOA at baseline. However, we found relatively high heterogeneity in model performance in leave-one-cohort-out cross-validation. The added value of hip morphology measurements on the discriminative performance was small compared to a model with baseline patient characteristics and RHOA grade alone. More efforts are recommended to reduce the heterogeneity between cohorts and increase model calibration. Ultimately, risk prediction models built with multicohort data could lead to an improved understanding of HOA and be a first step toward individualized HOA care.

ACKNOWLEDGMENTS

We would like to thank all participants of each of the cohort studies that are involved in World COACH. We gratefully acknowledge all international organizations that collaborated with the cohort studies in World COACH, as well as the OARSI for endorsing the World COACH consortium. The CHECK study was initiated by the Dutch Arthritis Society and performed within Erasmus Medical Center Rotterdam; Kennemer Gasthuis Haarlem; Leiden University Medical Center; Maastricht University Medical Center; Martini Hospital Groningen/Allied Health Care Center for Rheum. and Rehabilitation Groningen; Medical Spectrum Twente Enschede/Ziekenhuisgroep Twente Almelo; Reade, formerly Jan van Breemen Institute/VU Medical Center Amsterdam; St.Maartens-kliniek Nijmegen; University Medical Center Utrecht; and Wilhelmina Hospital Assen. We would like to thank all the participants of the Chingford Women Study, Professor Nigel Arden, Professor Tim Spector, Dr Deborah Hart, Gem Lawson, Maxine Daniels, and Alison Turner for their time and dedication and Arthritis Research UK for their funding support to the study and the Oxford NIHR Musculoskeletal Biomedical Research Unit for funding contributions. The Femoroacetabular impingement and

hip OsteoaRthritis Cohort was supported by the Australian National Health and Medical Research Council; Early Career Fellowship grant 1119971; Australian Postgraduate Scholarship. Support for data from the Johnston County Osteoarthritis Project was provided in part by the Center for Disease Control and Prevention (CDC) (grants U01DP006266 and U01DP003206); Association of Schools of Public Health/CDC (grants S043, S1734, S3486); and National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (P60AR30701, P60AR049465, P60AR064166, and P30AR072580). The MOST study was funded by the National Institute on Aging, NIH (grants AG19069 [Michael Nevitt, University of California, San Francisco], AG18820 [David Felson, Boston University], AG18947 [Cora Lewis, University of Alabama at Birmingham], and AG18832 [James Torner, University of Iowa]). The cohort, clinical data and image acquisitions used in these analyses were funded as the Osteoarthritis Initiative by the NIH through a Foundation for NIH public private partnership with GlaxoSmithKline, Merck & Co., Inc., Novartis Pharmaceuticals Corporation, and Pfizer. The Rotterdam Study is funded by Erasmus University Medical Center and Erasmus University, Rotterdam, The Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The Study of Osteoporotic Fractures is supported by NIH funding (grants R01-AG-005407, R01-AR-35582, R01-AR-35583, R01-AR-35584, R01-AG-005394, R01-AG-027574, and R01-AG-027576. The TASOAC study was supported by the National Health and Medical Research Council of Australia, Tasmanian Community Fund, Masonic Centenary Medical Research Foundation, Royal Hobart Hospital Research Foundation, and Arthritis Foundation of Australia.

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 Agricola 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. Fan Z, Yan L, Liu H, et al. The prevalence of hip osteoarthritis: a systematic review and meta-analysis. Arthritis Res Ther 2023;25(1):51.
2. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. Bone 2012; 51(2):278–288.
3. Shapira J, Chen JW, Bheem R, et al. Radiographic factors associated with hip osteoarthritis: a systematic review. J Hip Preserv Surg 2020; 7(1):4–13.
4. Agricola R, van Buuren MMA, Kemp JL, et al. Femoroacetabular impingement syndrome in middle-aged individuals is strongly associated with the development of hip osteoarthritis within 10-year follow-up: a prospective cohort study (CHECK). Br J Sports Med 2024; 58(18):1061–1067.
5. Melugin HP, Hale RF, Lee DR, et al. Risk factors for long-term hip osteoarthritis in patients with hip dysplasia without surgical intervention. J Hip Preserv Surg 2022;9(1):18–21.
6. Fischer CS, Kühn JP, Ittermann T, et al. What are the reference values and associated factors for center-edge angle and alpha angle? A population-based study. Clin Orthop Relat Res 2018;476(11):2249–2259.
7. Dijkstra HP, Mc Auliffe S, Ardern CL, et al; Young Athlete's Hip Research (YAHiR) Collaborative. Oxford consensus on primary cam morphology and femoroacetabular impingement syndrome: part 1-definitions, terminology, taxonomy and imaging outcomes. Br J Sports Med. 2022;57(6):325–341.
8. Appleyard T, Thomas MJ, Antcliff D, et al. Prediction models to estimate the future risk of osteoarthritis in the general population: a systematic review. Arthritis Care Res (Hoboken) 2023;75(7):1481–1493.
9. Steyerberg EW, Nieboer D, Debray TPA, et al. Assessment of heterogeneity in an individual participant data meta-analysis of prediction models: an overview and illustration. Stat Med 2019;38(22):4290–4309.
10. Wiemken TL, Kelley RR. Machine learning in epidemiology and health outcomes research. Annu Rev Public Health 2020;41(1):21–36.
11. Debray TPA, Collins GS, Riley RD, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data: TRIPOD-Cluster checklist. BMJ 2023;380:e071018.
12. van Buuren MMA, Riedstra NS, van den Berg MA, et al. Cohort profile: worldwide Collaboration on OsteoArthritis prediCtion for the Hip (World COACH) - an international consortium of prospective cohort studies with individual participant data on hip osteoarthritis. BMJ Open 2024;14(4):e077907.
13. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16(4):494–502.
14. Croft P, Cooper C, Wickham C, et al. Defining osteoarthritis of the hip for epidemiologic studies. Am J Epidemiol 1990;132(3):514–522.
15. Lindner C, Thiagarajah S, Wilkinson JM, et al; arcOGEN Consortium. Fully automatic segmentation of the proximal femur using random forest regression voting. IEEE Trans Med Imaging 2013;32(8):1462–1472.
16. Boel F, Riedstra NS, Tang J, et al. Reliability and agreement of manual and automated morphological radiographic hip measurements. Osteoarthr Cartil Open 2024;6(3):100510.
17. Boel F, de Vos-Jakobs S, Riedstra NS, et al. Automated radiographic hip morphology measurements: an open-access method. Osteoarthr Imaging 2024;4(2):100181.
18. Jacobsen S, Sonne-Holm S. Hip dysplasia: a significant risk factor for the development of hip osteoarthritis. A cross-sectional survey. Rheumatology (Oxford) 2005;44(2):211–218.
19. Lane NE, Lin P, Christiansen L, et al. Association of mild acetabular dysplasia with an increased risk of incident hip osteoarthritis in elderly white women: the study of osteoporotic fractures. Arthritis Rheum 2000;43(2):400–404.
20. Reid GD, Reid CG, Widmer N, et al. Femoroacetabular impingement syndrome: an underrecognized cause of hip pain and premature osteoarthritis? J Rheumatol 2010;37(7):1395–1404.
21. Reijman M, Hazes JM, Pols HA, et al. Acetabular dysplasia predicts incident osteoarthritis of the hip: the Rotterdam study. Arthritis Rheum 2005;52(3):787–793.
22. Runhaar J, van Berkel AC, Agricola R, et al. Risk factors and population-attributable fractions for incident hip osteoarthritis. HSS J 2023;19(4):407–412.
23. Saberi Hosnijeh F, Zuiderwijk ME, Versteeg M, et al. Cam deformity and acetabular dysplasia as risk factors for hip osteoarthritis. Arthritis Rheumatol 2017;69(1):86–93.
24. Ganz R, Parvizi J, Beck M, et al. Femoroacetabular impingement: a cause for osteoarthritis of the hip. Clin Orthop Relat Res 2003;417: 112–120.

25. Riedstra NS, Boel F, van Buuren MMA, et al. Acetabular dysplasia and the risk of developing hip osteoarthritis within 4-8 years; an individual participant data meta-analysis of 18,807 hips from the World COACH consortium. Osteoarthritis Cartilage. 2025;33(3):373–382.
26. Agricola R, Heijboer MP, Bierma-Zeinstra SM, et al. Cam impingement causes osteoarthritis of the hip: a nationwide prospective cohort study (CHECK). Ann Rheum Dis 2013;72(6):918–923.
27. van Klij P, Heerey J, Waarsing JH, et al. The prevalence of cam and pincer morphology and its association with development of hip osteoarthritis. J Orthop Sports Phys Ther 2018;48(4):230–238.
28. Chaganti RK, Lane NE. Risk factors for incident osteoarthritis of the hip and knee. Curr Rev Musculoskelet Med 2011;4(3):99–104.
29. Xiong C, Zhu K, Yu K, et al. Statistical modeling in biomedical research: longitudinal data analysis. In: Rao CR, Miller JP, Rao DC, eds. Essential Statistical Methods for Medical Statistics. Elsevier; 2011:235–268.
30. Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
31. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67(1).
32. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol 2022;75(1):25–36.
33. Pedregosa FVG, Gramfort A, Michel V, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–2830.
34. Casartelli NC, Maffiuletti NA, Valenzuela PL, et al. Is hip morphology a risk factor for developing hip osteoarthritis? A systematic review with meta-analysis. Osteoarthritis Cartilage 2021;29(9):1252–1264.
35. Riedstra NS, Vinge R, Herfkens J, et al. Acetabular dysplasia and the risk of developing hip osteoarthritis at 2,5,8, and 10 years follow-up in a prospective nationwide cohort study (CHECK). Semin Arthritis Rheum 2023;60:152194.
36. Saberi Hosnijeh F, Kavousi M, Boer CG, et al. Development of a prediction model for future risk of radiographic hip osteoarthritis. Osteoarthritis Cartilage 2018;26(4):540–546.
37. Damen J, Schiphof D, Wolde ST, et al. Inter-observer reliability for radiographic assessment of early osteoarthritis features: the CHECK (cohort hip and cohort knee) study. Osteoarthritis Cartilage 2014;22(7):969–974.
38. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353:i3140.
39. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 2002;137(7):598–602.
40. Costa ALF, Lopes SLPC. Highlighting the benefits and disadvantages of individual participant data meta-analysis in radiology. Radiol Imaging Cancer 2024;6(2):e240018.
41. Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. Clin Orthop Relat Res 2016;474(8):1886–1893.
42. Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. NPJ Digit Med 2022;5(1):171.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 501–511
DOI 10.1002/acr.25642

American College of Rheumatology
Empowering Rheumatology Professionals

Exploring Patients' Profiles Associated With the Resolution of Acute Calcium Pyrophosphate Arthritis Treated With Colchicine and Prednisone: Post Hoc Analysis of a Randomized Controlled Trial

Tristan Pascart,[1] Laurène Norberciak,[1] Pascal Richette,[2] Pierre Robinet,[1] Aurore Pacaud,[1] Gauthier Marchasson,[1] Thibault Rabin,[1] Hélène Luraschi,[1] Pierre Maciejasz,[3] Anne-France Georgel,[1] Augustin Latourte,[2] Hang-Korng Ea,[2] Sébastien Ottaviani,[4] Charlotte Jauffret,[1] and Vincent Ducoulombier[1]

Objective. The objective was to identify factors determining acute arthritis resolution and safety with colchicine and prednisone in acute calcium pyrophosphate (CPP) crystal arthritis.

Methods. We conducted a post hoc analysis of the COLCHICORT trial, which compared colchicine and prednisone for the treatment of acute CPP crystal arthritis, using a composite outcome of secondary endpoints of the primary analysis. Factors associated with the sustained arthritis resolution with prednisone or colchicine treatment on day three, and the occurrence of gastrointestinal adverse events (AEs) with colchicine, were examined. Two different machine-learning approaches were used to identify independent factors associated with the efficacy outcome: multiple logistic regressions and a decision tree were used.

Results. In total, 43 of 89 (48.3%) of participants were considered definite responders. In univariable analysis, definitive responders were more often hospitalized for stroke (9 of 43; $P = 0.04$), had an age ≥80 years old (39 of 43; $P = 0.045$), and were dyslipidemic (25 of 43; $P = 0.03$), whereas poor responders were more commonly hospitalized to manage the acute arthritis episode (22 of 43; $P < 0.001$) In multiple logistic regression, acute arthritis of the wrists (odds ratio [OR] 4.06, 95% confidence interval [CI] 1.21–15.50.85) were associated with arthritis resolution on day three, whereas randomization in the colchicine arm (OR 0.31, 95% CI 0.11–0.83) and diuretic use (OR 0.23, 95% CI 0.097–0.95) were associated with a poor treatment response. Hospital admission for acute arthritis, C-reactive protein levels, and estimated glomerular filtration rate were decision tree nodes selected as crucial for predicting definitive flare resolution. Three candidate variables were identified in the multiple logistic regression model explaining the occurrence of gastrointestinal AEs with colchicine: male sex (OR 0.33, 95% CI 0.07–1.29) and diabetes (OR 0.24, 95% CI 0.030–1.24) seemed protective, whereas statin use (OR 3.54, 95% CI 0.83–18.82) seemed associated with their occurrence.

Conclusion. This study identified factors associated with the definitive response to colchicine and prednisone treatment in acute CPP crystal arthritis. Colchicine was associated with poorer efficacy and was impacted by a dual effect on its safety and efficacy by medications and associated conditions.

INTRODUCTION

Calcium pyrophosphate deposition (CPPD) disease occurs as a consequence of the inflammatory response and cartilage damage because of the pathologic presence of calcium pyrophosphate (CPP) crystals inside joints.[1–7] It is an umbrella term for different acute and chronic inflammatory and noninflammatory phenotypes, which often coexist.[7–9] Acute CPP crystal arthritis is

Supported by the French Inter-regional Hospital Clinical Research Programme (2016).

[1]Tristan Pascart, MD, PhD, Laurène Norberciak, BsC, Pierre Robinet, MD, Aurore Pacaud, MD, Gauthier Marchasson, MD, Thibault Rabin, MD, Hélène Luraschi, MD, Anne-France Georgel, PhD, Charlotte Jauffret, MD, Vincent Ducoulombier, MD: Lille Catholic University, Saint-Philibert Hospital, Lille, France; [2]Pascal Richette, MD, PhD, Augustin Latourte, MD, PhD, Hang-Korng Ea, MD, PhD: Université de Paris, INSERM, UMR-S 1132 BIOSCAR AP-HP, Lariboisière Hospital, Paris, France; [3]Pierre Maciejasz, MD: Lille Catholic University, Saint-Vincent-de-Paul Hospital, Lille, France; [4]Sébastien Ottaviani, MD: Université Paris Cité, Hôpital Bichat-Claude Bernard, Paris, France.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25642).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25642.

Address correspondence via email to Tristan Pascart, MD, PhD, at pascart.tristan@ghic1.net.

Submitted for publication November 28, 2024; accepted in revised form August 4, 2025.

SIGNIFICANCE & INNOVATIONS

- This post hoc analysis of the first trial performed in acute CPP crystal arthritis identified factors (demographic, acute arthritis characteristics, associated conditions [including underlying osteoarthritis], and medications) predictive of arthritis resolution with prednisone and colchicine.
- Using for this post hoc analysis a different primary endpoint than the one used for the primary analysis of the COLCHICORT trial, prednisone provides more acute CPP crystal arthritis resolution after two days of treatment than colchicine.
- Associated conditions and medications have a dual inverse effect on colchicine's efficacy and safety.

the most frequently recognized phenotype and is characterized by severe joint pain and swelling, like gout flares, commonly affecting the knees, wrists, and ankles, and lasting for several days to weeks.[10–13] Acute CPP crystal arthritis is the most common cause of acute arthritis in older adults, and frequently occurs during in-hospital stays, and its management is at the crossroads of internal medicine, geriatrics, rheumatology, and general practice.[1,12] The objective of the management of these acute flares of CPP crystal arthritis is to obtain rapid and sustained pain relief, mainly by controlling joint inflammation, without producing adverse events (AEs) in the fragile target population of older people.[1] Anti-inflammatory strategies for the treatment of acute CPP crystal arthritis can rely on colchicine, prednisone, nonsteroidal anti-inflammatory drugs, intraarticular glucocorticoid injections, and, in some difficult cases, targeted therapies, mainly interleukin-1 (IL-1) or IL-6 inhibitors.[1] No guidance has yet been provided on which drug to choose among these options for the treatment of acute CPP crystal arthritis, mainly because of a lack of comparative evidence of efficacy and safety but also because no factors predicting treatment response or safety have been clearly identified so far.[1,3,14] In particular, how the specifics of older people, including decreased renal function, age, and comedications, impact the efficacy and safety of drugs targeting inflammation, remains unknown and adds to the difficulties of managing acute CPP crystal arthritis for physicians.

The COLCHICORT trial was the first-ever completed randomized controlled trial studying acute CPP crystal arthritis and compared the efficacy after the first day of treatment and safety of a two-day regimen of colchicine or prednisone.[11] The drugs provided equivalent pain relief on day two, with a safety profile favoring prednisone because of the occurrence of gastrointestinal AEs, mainly diarrhea, in 22% of participants in the colchicine arm. The first analyses provided hints that colchicine had better response rates when administered within 12 hours of arthritis onset (odds ratio [OR] 4.0, 95% confidence interval [CI] 0.7–21.0), that statin use may increase the risk of colchicine-induced diarrhea (OR 3.0, 95% CI 0.9–10.6), and that prednisone use may require less additional treatments to obtain definitive flare resolution.[11] Pain relief after the second day of treatment was a secondary outcome and did not significantly differ between treatment groups. Patients with colchicine seemed to require more additional anti-inflammatory drugs in the follow-up than those randomized in the prednisone group. The objective of this post hoc analysis was to identify factors associated with acute CPP crystal arthritis resolution sustained after day three with treatments with colchicine and prednisone and their safety, to help choose between available drugs.

PATIENTS AND METHODS

Participants and design. We conducted a post hoc analysis of the COLCHICORT equivalence randomized controlled trial, which compared colchicine and prednisone for the treatment of acute CPP crystal arthritis, and showed equivalence between the two drugs in obtaining pain relief at 24 hours (day two) (NCT03128905).[11] Detailed methods and results of the full trial have been reported previously.[11] In short, this was a 7-day, open-label equivalence trial with participants randomized 1:1 to colchicine 1.5 mg on day one and 1 mg on day two or prednisone 30 mg on day one and two. Systematic analgesia was given in the first 24 hours, with 50 mg tramadol and 1 g paracetamol every 8 hours. After the second dose of the experimental drugs on day two, investigators were free to decide how to manage the episode until the last assessment on day eight, including using additional doses of colchicine, prednisone, and intraarticular glucocorticoid injections. The primary endpoint was the change in visual analog scale (VAS) pain (0–100) from baseline on day two (24 hours). Ninety-five patients were included in the analysis of the per protocol (PP) population. Secondary endpoints included the change in VAS pain from baseline to day three (48 hours) and the need for additional analgesics and anti-inflammatory drugs (additional colchicine/prednisone doses or intraarticular glucocorticoid injections).

All participants presented with acute CPP crystal arthritis with a symptom duration of <36 hours and pain intensity at ≥40 mm (/100 mm) on a VAS for the most painful joint. Diagnosis was ascertained by identifying CPP crystals on synovial fluid aspiration or with imaging evidence of CPPD,[15,16] and all of them retrospectively met the newly developed American College of Rheumatology/EULAR classification criteria for CPPD disease.[10]

The trial was sponsored by the 2016 Northwest GIRCI French Inter-regional Hospital Clinical Research Programme (PHRC no. 16.31). The trial protocol was reviewed and approved by a French national institutional review board (CPP Nord-Ouest number 17/43, registration number 2016, ANSM number 170356A-21). All patients provided written informed consent to participate in the study and the study complied with the French Data Protection Authority (CNIL)'s MR001 reference methodology.

Outcomes of treatment response and safety. In this post hoc analysis, the main efficacy outcome was a composite outcome of secondary endpoints of the primary analysis and was defined as a change on the pain VAS from baseline of ≥50% and/or a pain VAS <40/100 by day three (after the full two days of the experimental phase) and no further use of colchicine, prednisone, or any other systemic anti-inflammatory agent and no intraarticular glucocorticoid injection.

The safety outcome examined was the occurrence of any gastrointestinal AE (diarrhea, abdominal pain, nausea, or vomiting) in participants randomized in the colchicine arm. No other AEs were sufficiently frequent to examine risk factors.

Variables. In the current analysis, a number of baseline variables were examined, including (1) general demographic and clinical variables such as age, sex, body mass index, blood pressure, presence of comorbidities, and estimated glomerular filtration rate (eGFR); (2) clinical and biologic characteristics of the acute CPP crystal arthritis (affected joint(s), pain VAS, duration since onset, cause of hospital admission, arthrocentesis performed or not, prior episode of acute arthritis, C-reactive protein (CRP) level, neutrophil and monocyte counts); (3) conventional radiography evidence of CPPD and osteoarthritis (OA) in specific and symptomatic sites; (4) current drugs; and (5) randomized group (colchicine vs prednisone).

Statistical analysis. In the absence of any prior data orienting the selection of baseline characteristics that may or may not have been associated with treatment response, we chose to use an unoriented automatic selection of variables, which is otherwise avoided.[17] These post hoc analyses had not been prespecified in the original protocol. Data exploration and statistical analysis were conducted using the R programming language (R version 4.3.0).

Descriptive analysis provided means and standard deviations for normally distributed quantitative variables, and medians and interquartile ranges (IQRs) for nonnormally distributed quantitative variables. The normality of the data was verified using the Shapiro-Wilk test. Counts and frequencies are presented for qualitative variables. Some continuous variables were dichotomized. The best threshold value for these variables was chosen from the receiver operating characteristic curve according to the Youden index (maximum sum of sensitivities [Se] + specificities [Sp]). Univariable analyses for the efficacy outcome, first combining both treatments and then for each treatment arm individually, and for the safety outcome of colchicine were performed using the chi-squared or Fisher exact tests for qualitative data according to the Cochran rule and with Student's t tests or Wilcoxon-Mann-Whitney tests for quantitative data according to data normality using the Shapiro-Wilk test. Unadjusted ORs and their 95% CIs were calculated. These univariable analyses were performed to inform the multivariable analysis and were therefore not adjusted for multiple comparisons. Two complementary multivariable models, a multiple logistic regression model with stepwise backward elimination and a decision classification and regression tree (CART) model, were fitted for the efficacy outcome of the treatments combined. An insufficient sample size in the colchicine group alone prevented a consistent decision tree for the occurrence of any gastrointestinal AE with colchicine and multivariable models for the efficacy outcome of individual treatments.

Variables with a between-group difference associated with a P value of <0.20 or with a ≥10-point difference, and with <10% missing values, and with each binary condition present in >10% of the sample, were entered into a multiple logistic regression model to which a stepwise backward elimination based on Akaike information criterion (AIC) was applied to obtain a reduced model. Multivariable ORs and their 95% CI of the reduced model were calculated.

Additionally, the predictive performance of the "traditional" logistic regression model was supplemented by a decision tree approach using the CART algorithm with the rpart package in R, based on the Gini impurity index.[18] The choice of that alternative is justified by many advantages, namely nonparametric modeling, interpretability, and visualization. The CART algorithm was used to search for the split on each variable to partition the data into two groups depending on a binary response (treatment response and no treatment response). The CART method iteratively split data into subgroups (nodes) using a hierarchical strategy. That process yields a tree, and the subgroups in the final partition (the leaves) are associated with predictions given by the majority response within each leaf. The CART method identifies the best overall split (partition) optimizing a loss function based on the impurity criterion and under certain constraints (eg minimal size of leaves, etc). Similarly to the AIC for variable selection in logistic regression, pruning was then applied to generate a parsimonious tree model using the "cost complexity" method (ie, eliminating leaves that do not significantly increase accuracy in order to prevent overfitting). Within each (intermediate and final) node, the tree plot provides classes of patients, labeled with the dominant response modality, the percentage of participants classified as good responders, and the percentage of the sample they represent.

For both multivariable methods, all patients were considered in the training set without test set, and unbiased logistic regression and tree performances were calculated with leave-one-out cross validation (LOOCV); the error rate (misclassification) and Se and Sp (based on a 50% probability cutoff) were calculated. Statistical significance was set at $P < 0.05$.

Patient consent for publication. All patients provided written informed consent to participate in the study and the study complied with the French Data Protection Authority (CNIL)'s MR001 reference methodology.

Table 1. Patients' characteristics according to the combined response to treatment with colchicine and prednisone (change in pain VAS from baseline of ≥50% and/or pain VAS <40/100 by day three (48 hours) and no further use of anti-inflammatory agents)*

Baseline characteristics	Per protocol population (n = 95)	No definitive flare resolution (48 h) (n = 46)	Definitive flare resolution (48 h) (n = 43)	*P* value
Anthropometry				
Age, median (IQR), years	88 (82–91)	88 (79–91)	87 (84.5–90.5)	0.63
Age, >80 years, n (%)	72 (80.9)	33 (71.7)	39 (90.7)	0.045
Male, n (%)	26 (27.4)	9 (19.6)	13 (30.2)	0.36
Body mass index, mean (SD)	25.7 (5.47)	26.4 (6.3)	25 (4.6)	0.24
Body mass index >25, n (%)	48 (50.5)	25 (54.3)	21 (48.8)	0.76
SAP >124 mm Hg, n (%)	58 (61.7)	31 (68.9)	23 (53.5)	0.21
Comorbidities				
High blood pressure, n (%)	76 (80)	37 (80.4)	35 (81.4)	1
Type 2 diabetes, n (%)	24 (25.3)	12 (26.1)	12 (27.9)	1
Dyslipidemia, n (%)	44 (46.3)	15 (32.6)	25 (58.1)	0.027
Current treatments, n (%)				
Statins	28 (29.5)	12 (26.1)	15 (34.9)	0.5
Antidiabetics	16 (16.8)	9 (19.6)	7 (16.3)	0.9
Diuretics	28 (29.5)	17 (37)	10 (23.3)	0.24
ACE/RAA2 inhibitors	26 (27.4)	11 (23.9)	12 (27.9)	0.85
Beta blockers	48 (50.5)	23 (50)	25 (58.1)	0.58
Calcium channel inhibitors	21 (22.1)	8 (17.4)	12 (27.9)	0.35
Proton pump inhibitors	32 (33.7)	15 (32.6)	16 (37.2)	0.82
Anticoagulants	65 (68.4)	30 (65.22)	32 (74.42)	0.48
Characteristics of the acute CPP crystal arthritis episode				
Initial cause of hospital admission, n (%)				
Acute arthritis	26 (27.4)	22 (47.8)	4 (9.3)	0.00017
Trauma	28 (29.5)	12 (26.1)	14 (32.6)	0.66
Stroke	11 (11.6)	2 (4.3)	9 (20.9)	0.04
General deterioration in health	12 (12.6)	4 (8.7)	6 (14)	0.51
Infection dyspnea	7 (7.4)	3 (6.5)	3 (7)	1
Prior episode of acute CPP crystal arthritis, n (%)	25 (26.3)	10 (21.7)	13 (30.2)	0.5
Painful joint, n (%)				
Shoulder	7 (7.4)	3 (6.5)	3 (7)	1
Elbow	4 (4.2)	3 (6.5)	1 (2.3)	0.62
Hip	0 (0)	0 (0)	0 (0)	—
Wrist	21 (22.1)	6 (13)	13 (30.2)	0.086
Knee	51 (53.7)	29 (63)	19 (44.2)	0.12
Ankle	20 (21.1)	12 (26.1)	7 (16.3)	0.38
Cervical spine (CDS)	8 (8.4)	5 (10.9)	3 (7)	0.71
Metacarpophalangeal joint	5 (5.3)	2 (4.3)	3 (7)	0.67
Pubic symphysis	0 (0)	0 (0)	0 (0)	—
Other	3 (3.2)	2 (4.3)	1 (2.3)	1
Baseline pain VAS >70/100 (mm), n (%)	41 (46.1)	22 (47.8)	19 (44.2)	0.89
More than one painful joint, n (%)	18 (18.9)	10 (21.7)	7 (16.3)	0.7
Duration of the acute arthritis episode <12 hours, n (%)	32 (34)	14 (31.1)	14 (32.6)	1
Arthrocentesis performed, n (%)	43 (45.3)	22 (47.8)	19 (44.2)	0.9
Randomization arm, n (%)				
Prednisone	46 (48.4)	18 (39.1)	25 (58.1)	0.11
Colchicine	49 (51.6)	28 (60.9)	18 (41.9)	
Laboratory results				
C-reactive protein, median (IQR), mg/L	80 (39–143.5]	91.5 (56.8–148.2)	75 (32–132)	0.23
eGFR < 60 mL/min/1.73 m^2, n (%)	41 (48.8)	20 (45.5)	21 (52.5)	0.67
Hypomagnesemia	6 (9.5)	1 (3.4)	5 (17.2)	0.19
Neutrophils >10,000/mm^3, n (%)	14 (17.1)	6 (14)	8 (20.5)	0.62
Monocytes >1,000/mm^3, n (%)	39 (47.6)	19 (44.2)	20 (51.3)	0,67
Imaging, n (%)				
CPPD in the index joint	84 (94.4)	42 (95.5)	36 (92.3)	0.66
Wrist CPPD	58 (73.4)	29 (72.5)	28 (80)	0.63
Knee CPPD	49 (65.3)	21 (61.8)	25 (69.4)	0.67
Pubic symphysis CPPD	34 (41.5)	14 (35.9)	18 (48.6)	0.37
Osteoarthritis in the index joint	51 (60.7)	25 (59.5)	23 (63.9)	0.87
Metacarpophalangeal joint osteoarthritis	26 (32.9)	14 (35)	12 (34.3)	1

(Continued)

Table 1. (*Cont'd*)

Baseline characteristics	Per protocol population (n = 95)	No definitive flare resolution (48 h) (n = 46)	Definitive flare resolution (48 h) (n = 43)	*P* value
Scaphotrapeziotrapezoidal joint osteoarthritis	28 (35.4)	15 (37.5)	12 (34.3)	0.96
Wrist osteoarthritis	29 (36.7)	12 (30)	17 (48.6)	0.16
Knee osteoarthritis	34 (45.3)	14 (41.2)	20 (55.6)	0.34

* Univariable comparisons between responders and nonresponders were performed. ACE, angiotensin-converting enzyme; CDS, crowned dens syndrome; CI, confidence interval; CPP, calcium pyrophosphate; CPPD, CPP deposition; eGFR, estimated glomerular filtration rate; IQR, interquartile range; RAA2, renin-angiotensin antagonist; SAP, systolic arterial pressure; VAS, visual analog scale.

Ethics approval. The trial protocol was reviewed and approved by a French national IRB (CPP Nord-Ouest number 17/43, registration number 2016, ANSM number 170356A-21).

RESULTS

Participants. The PP population of the COLCHICORT trial included 95 patients, 46 of whom were treated with prednisone and 49 of whom were treated with colchicine. A total of 89 participants had pain assessment on day two and were included in this analysis, 43 from the prednisone group and 46 from the colchicine group. The characteristics of the participants are described in Table 1. Briefly, the median age was 88 years (IQR 82–91; 26 participants (27.4%) were male, 24 participants (25%) had diabetes, and 76 participants (80%) had hypertension. The median eGFR was 62.1 mL/min per 1.73 m^2 (IQR 45.2–81.5). Acute CCP crystal arthritis with a first-ever flare was reported in 70 participants (73.7%), and the acute arthritis affected mainly the knees in 46 participants (48%), the wrists in 19 participants (20%), and the ankles in 12 participants (13%) (Table 1).

Predictive factors of sustained arthritis resolution after day three with colchicine or prednisone. A total of 31 participants (72.1%) in the prednisone group and 30 participants (65.2%) in the colchicine group were considered good treatment responders on day three (P = 0.64). Overall, 51 of 89 patients (57.3%) did not require any additional anti-inflammatory treatments (colchicine/prednisone/intraarticular glucocorticoid injections) after 48 hours, and the flare was considered definitively resolved on day three. In total, 43 of 89 (48.3%) of participants were considered definite responders. In univariable analysis, definitive responders were more often hospitalized for stroke (9 of 43; P = 0.04), had an age ≥80 years old (39 of 43; P = 0.045) and were dyslipidemic (25 of 43; P = 0.03), whereas poor responders were more commonly hospitalized to manage the acute arthritis episode (22 of 43; P < 0.001; Table 1).

The 89 participants with complete data were included in the logistic regression model, which examined all relevant variables in the univariable analysis except eight variables, which had >10% of missing data: magnesium levels, calcium levels, transferrin levels, OA of the symptomatic joint, and conventional radiography evidence of CPPD and OA at the knees and wrists. History of stroke, age ≥80 years old, and admission for acute arthritis, all significantly associated with treatment response in univariable analysis, could not be included in the model because one binary condition represented <10% of the population. The model retained five explanatory variables, three of which were significantly associated with treatment response: acute arthritis affecting wrists (OR 4.06, 95% CI 1.21–15.50) was associated with definitive flare resolution at 48 hours in the reduced model, whereas randomization in the colchicine arm (OR 0.31, 95% CI 0.11–0.83) and diuretic use (OR 0.32, 95% CI 0.097–0.95) were negatively associated with definitive flare resolution (Figure 1). The error rate was 36% in the training set and 41.6% in the LOOCV, associated with an Se of 48.8% and Sp of 67.4%.

The pruned decision tree obtained with the CART algorithm integrated threshold values for CRP and eGFR and being hospitalized for the acute CPP arthritis or not (Figure 2). The best-case scenario was for participants hospitalized for other reasons than the acute CPP arthritis episode, with CRP levels of ≥88 mg/L and eGFR <43 mL/min/1.73 m^2, which represented 9% of patients, and 86% had a definitive flare resolution after 48 hours of treatment. The worst-case scenario was for participants hospitalized because of their acute CPP arthritis episode. They represented 31% of participants and were rapid responders in 17% of cases. The rate of participants wrongly classified by the tree was 21.8% in the training set and 26.9% in the LOOCV, associated with an Se of 64.9% and Sp of 80.5%. This tree correctly classifies two of three definitive responses. The predictive performance of the decision tree was better than that of the regression model, although a little less sensitive.

Factors associated with sustained arthritis resolution after day three with colchicine and prednisone individually. The unadjusted univariable analysis of response to each treatment individually is presented in Supplementary Table 1. The context of the hospital admission for the acute arthritis was associated with a poor response to both drugs (OR 0.05, 95% CI 0.00–0.30, for colchicine; and OR 0.21, 95% CI 0.04–0.93, for prednisone), whereas hospital admission for stroke was associated with a good treatment response in all patients treated with prednisone but did not affect response to colchicine

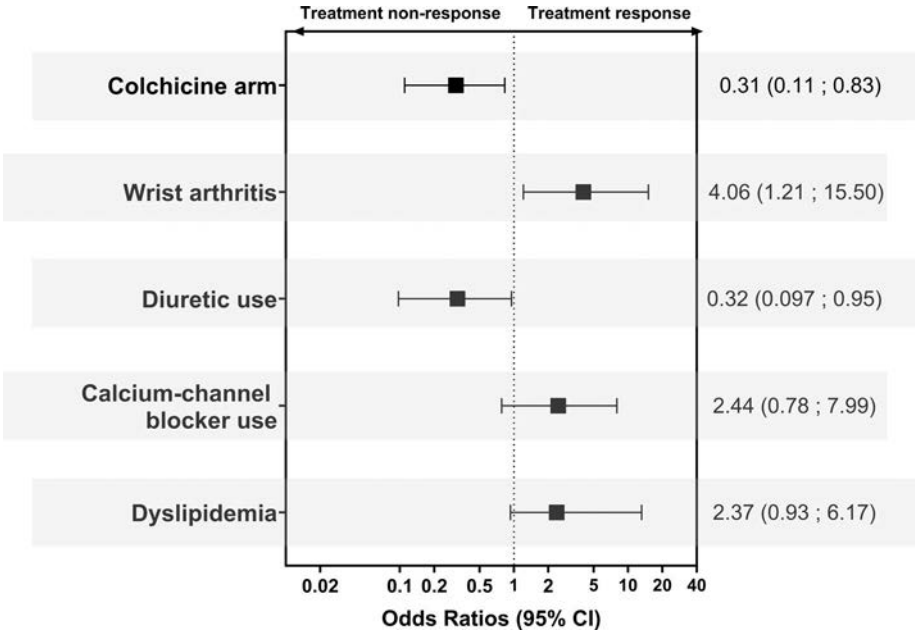

Figure 1. Forest plot of independent factors selected in the reduced multiple logistic regression explaining definitive flare resolution by day three defined by a change in pain VAS from baseline of ≥50% and/or pain VAS <40/100 by day three (48 hours) and no further use of anti-inflammatory agents. Odds ratios and their 95% CI of variables selected in the reduced model are presented. CI, confidence interval; VAS, visual analog scale.

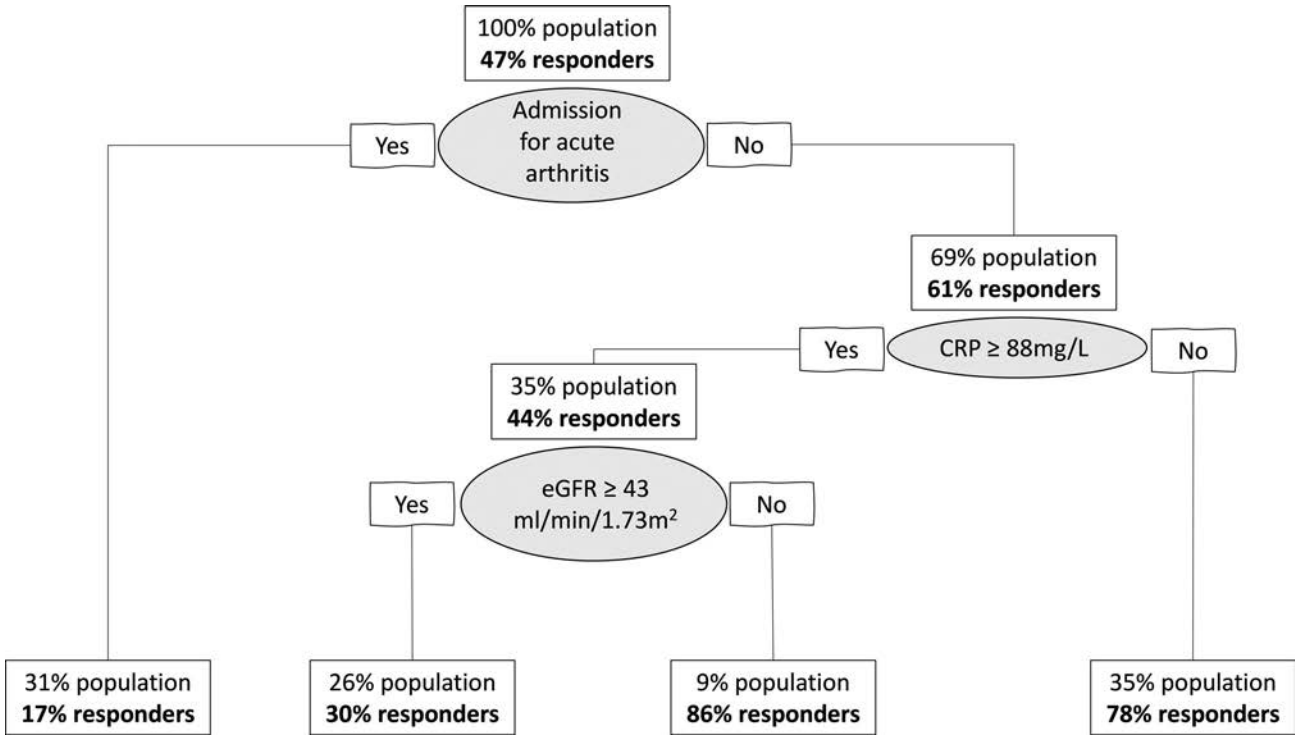

Figure 2. Illustration of the decision tree model generated by CART algorithm for definitive flare resolution by day three defined by a change in pain VAS from baseline of ≥50% and/or pain VAS <40/100 by day three (48 hours) and no further use of anti-inflammatory agents. The percentage of treatment responders and the percentage of patients from the overall sample are shown in each box representing a tree node, after the name of the dominant class in each node. At each split, an observation goes to the left branch if the condition is satisfied and goes on the right otherwise. The rate of participants wrongly classified by the tree was 21.8% in the training set and 26.9% in the leave-one-out crossvalidation. CART, classification and regression tree; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; VAS, visual analog scale.

Table 2. Patients' characteristics according to the occurrence of gastrointestinal adverse events in the colchicine group*

Baseline characteristics	Colchicine arm (n = 55)	No gastrointestinal adverse event (n = 35)	Gastrointestinal adverse events (n = 20)	*P* value
Anthropometry				
Age >80, median (IQR), years	41 (74.6)	25 (71.4)	16 (80)	0.70
Male, n (%)	17 (30.9)	14 (40)	3 (15)	0.10
Body mass index ≥25, n (%)	26 (47.3)	17 (48.6)	9 (45)	1
Comorbidities				
High blood pressure, n (%)	45 (81.8)	26 (74.3)	19 (95)	0.075
Type 2 diabetes, n (%)	13 (23.6)	10 (28.6)	3 (15)	0.33
Dyslipidemia, n (%)	24 (43.6)	14 (40)	10 (50)	0.66
Current treatments, n (%)				
Statins	16 (29.1)	8 (22.9)	8 (40)	0.3
Anti-diabetics	8 (14.5)	5 (14.3)	3 (15)	1
Diuretics	10 (18.2)	3 (8.6)	7 (35)	0.026
ACE/RAA2 inhibitors	15 (27.3)	10 (28.6)	5 (25)	1
Beta blockers	27 (49.1)	19 (54.3)	8 (40)	0.46
Calcium channel inhibitors	14 (25.5)	9 (25.7)	5 (25)	1
Proton pump inhibitors	15 (27.3)	10 (28.6)	5 (25)	1
Antibiotics	9 (16.4)	6 (17.1)	3 (15)	1
Laxatives	17 (30.9)	10 (28.6)	7 (35)	0.85
Antinausea	3 (5.5)	2 (5.7)	1 (5)	1
Anticoagulants	42 (76.4)	25 (71.4)	17 (85)	0.33
Characteristics of the acute CPP crystal arthritis episode				
Initial cause of hospital admission, n (%)				
Acute arthritis	16 (29.1)	11 (31.4)	5 (25)	0.84
Trauma	12 (21.8)	8 (22.9)	4 (20)	1
Stroke	5 (9.1)	4 (11.4)	1 (5)	0.64
General deterioration in health	5 (9.1)	2 (5.7)	3 (15)	0.34
Infection, dyspnea	4 (7.3)	2 (5.7)	2 (10)	0.62
Prior episode of acute CPP crystal arthritis	15 (27.3)	9 (25.7)	6 (30)	0.98
Baseline symptoms				
Digestive disorders	3 (5.5)	0 (0)	3 (15)	0.043
Anxiety/insomnia/delirium	7 (12.7)	5 (14.3)	2 (10)	1
Baseline pain VAS ≥70/100, mm	24 (43.6)	16 (45.7)	8 (40)	0.90
Imaging, CPPD in the index joint, n (%)	47 (95.9)	29 (93.5)	18 (100)	0.53
Laboratory results				
eGFR ≤-à mL/min/1.73 m^2, n (%)	22 (43.1)	12 (35.3)	10 (58.8)	0.19
Serum magnesium <0.66 mmol/L, n (%)	3 (8.3)	2 (8.3)	1 (8.3)	1

* Univariable comparisons between responders and nonresponders were performed. ACE, angiotensin-converting enzyme; CPP, calcium pyrophosphate; CPPD, CPP deposition; eGFR, estimated glomerular filtration rate; IQR, interquartile range; RAA2, renin-angiotensin antagonist; VAS, visual analog scale.

(OR 2.60, 95% CI 0.39–21.48). Coprescribed diuretics tended to be associated with a lower response to both colchicine (OR 0.18, 95% CI 0.01–1.13) and prednisone, to a lesser extent, (OR 0.45, 95% CI 0.13–1.53) but did not reach statistical significance. Participants treated with antidiabetic drugs seemed less likely to be responders to prednisone (OR 0.35, 95% CI 0.06–1.68) than those treated with colchicine (OR 1.71, 95% CI 0.36–8.33), and a diagnosis of type 2 diabetes tended to favor a response to colchicine rather than to prednisone. Calcium channel blockers were associated with a good response to colchicine (OR 3.68, 95% CI 0.99–15.01) but did not affect response to prednisone (OR 0.95, 95% CI 0.18–5.43). Responders to colchicine had a lower eGFR compared to nonresponders, (53.3 mL/min/1.73 m^2 [IQR 36.8–82] vs 73.7 mL/min/1.73 m^2 [IQR 56.2–86.4]; $P = 0.07$), whereas renal function was consistent between responders and nonresponders of the prednisone group ($P = 0.54$). Acute CPP crystal arthritis in the knees and ankles had lower response rates with colchicine (OR 0.28, 95% CI 0.08–0.94 and OR 0.18, 95% CI 0.01–1.13, respectively), whereas the efficacy of prednisone did not seem to be affected (OR 0.69, 95% CI 0.20–2.34 and OR 0.82, 95% CI 0.20–3.39, respectively). Although the presence of OA in any joints did not affect the response to colchicine, OA in the knees (OR 20.43, 95% CI 3.05–415.00) and wrist (OR 6.50, 95% CI 1.52–35.60) was associated with a good treatment response to prednisone. Responders in the prednisone group more often had OA in the index joint (64% vs 42%, non significant [NS]) contrary to the colchicine group (64% vs 72%, NS). Imaging evidence of CPPD in the index joint or any other joint did not affect treatment for either of the two drugs.

Because of an insufficient sample size for each treatment taken individually, no multivariable analysis was performed to examine separate associations with treatment response.

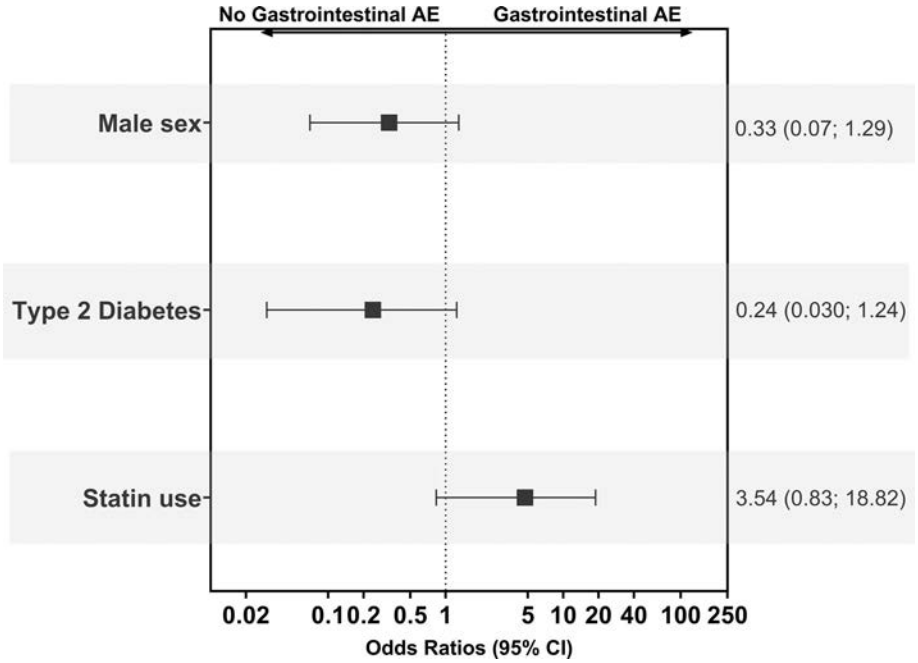

Figure 3. Forest plot of independent factors selected in the reduced multiple logistic regression explaining the occurrence of at least one gastrointestinal adverse event (diarrhea, abdominal pain, nausea, or vomiting) with colchicine. ORs and their 95% CIs of variables selected in the reduced model are presented. CI, confidence interval; OR, odds ratio.

Factors associated with gastrointestinal AEs with colchicine. Among the participants randomized in the colchicine arm, 20 of 55 (36.4%) experienced at least one gastrointestinal AE. In the univariable analysis, already having digestive disorders at baseline (3 of 20; $P = 0.04$) and treatment with diuretics (7 of 20; $P = 0.03$) were the only variables significantly associated with gastrointestinal AEs (Table 2). The following variables were included in the multiple logistic regression model, which included 49 participants with complete data: sex, diabetes, dyslipidemia, statin use, diuretic use, beta-blocker use, and anticoagulant use. Three explanatory variables were eventually retained in the model, none of which were statistically significantly associated with the occurrence gastrointestinal AEs related to colchicine: male sex (OR 0.33, 95% CI 0.07–1.29) and diabetes (OR 0.24, 95% CI 0.030–1.24) seemed protective, whereas statin use (OR 3.54, 95% CI 0.83–18.82) seemed to be associated with the AE (Figure 3).

DISCUSSION

This post hoc analysis provides new factors associated with a new definition of treatment response after a two-day regimen of colchicine or prednisone for acute CPP crystal arthritis. These factors include characteristics of the acute flare (site and inflammatory biomarkers), of the patients (age, gender, and renal function), evidence of OA, and current drugs. Some of these factors were shared by both treatments, and some were treatment-specific. Contrary to the analysis of the primary efficacy outcome of rapid pain relief on day two in the COLCHICORT trial,[11] the treatment arm did influence sustained arthritis resolution after day three, favoring prednisone in the logistic regression model. The study also provides new factors associated with the onset of gastrointestinal AEs with colchicine treatment. In most cases, these factors were also those associated with a good response.

The study identified new factors associated with sustained treatment response to colchicine and prednisone for the treatment of acute CPP crystal arthritis. The first analyses of the COLCHICORT trial showed that treatment administered within 12 hours after flare onset was associated with a better response to colchicine, similarly as it had been shown in the AGREE trial in gout flares but did not impact the response to prednisone.[11,19] However, the rapidity of colchicine administration did not seem to impact flare resolution on day three in this new analysis. This post hoc analysis provides new profiles of patients more likely to be treatment responders. Overall, patients with nonsevere CPP crystal arthritis affecting the wrist rather than the knee, with lower levels of markers of inflammation and pain intensity and impaired renal function, had a higher chance of having their pain sufficiently relieved by day three without any need for additional anti-inflammatory agents. Flares affecting knees and ankles seemed to be less responsive to colchicine, whereas the type of joint involved did not affect prednisone efficacy. In addition, the presence of OA (in wrists and knees) was highly associated with a good treatment response to prednisone but did not affect response to colchicine. This differential response between the

two drugs in case of underlying OA is consistent with the negative results of colchicine in chronic symptoms of OA, whereas prednisolone has been shown to relieve symptoms of hand OA.[20–23] Patients treated with calcium channel blockers seemed to be better responders, whereas those treated with diuretics more rarely had a sustained flare resolution by day three. Patients hospitalized to manage their acute CPP crystal arthritis seemed to have more difficult-to-treat episodes, whichever treatment was used. Although the COLCHICORT trial demonstrated the equivalence between colchicine and prednisone in achieving rapid pain relief by day two,[11] this post hoc analysis suggests that prednisone more often provides pain relief after the 2-day treatment regimen, without further need for other anti-inflammatory agents. This finding was suggested by the multiple logistic regression model, but the colchicine arm was not selected by the decision tree as being significant, imposing caution with this interpretation. Combined with the relatively better safety profile identified in the trial analyses first published,[11] these elements suggesting prednisone's superior efficacy to colchicine further supports its position as the first-line option for the treatment of acute CPP crystal arthritis, which can be adjusted according to the predictive factors that we identified, particularly in case of type 2 diabetes, in which case colchicine seems a better option.

Decision trees are not commonly presented for decision-making from trial results but provide intuitive guidance for personalized medicine according to patient profile clusters.[24–27] In our analyses, the CART analysis provided better results of sensitivity and specificity in crossvalidation than the logistic regression model did. The tree nodes determined which factors were decisive in each subpopulation obtained by the previous nodes to predict the treatment response. The multiple discretization is one of the main features of classification trees because the tree determines the most discriminating thresholds for quantitative variables in each node subpopulation. The decision tree clustering profiles of treatment further illustrated the importance of CRP levels, renal function, and initial reason for hospital admission to predict treatment response. The tree's performances had decent combined Se and Sp, with a one in four chance of misclassification, but will need to be tested in other populations to confirm that these results can be extrapolated.

Drug–drug interaction is a well-known safety issue when considering the use of colchicine, mainly mediated by colchicine being metabolized both by cytochrome P450 3A4 (CYP 3A4) and *P*-glycoprotein (*P*-gp) located in the enterocytes and bile ducts.[28] The only firm available data so far on the drug–drug interactions of colchicine is with the strongest inhibitors of CYP 3A4 (macrolide antibiotics and ketoconazole) and of *P*-gp (cyclosporine, tacrolimus, verapamil, and diltiazem), which either mildly or strongly influence colchicine exposure in a study of pharmacokinetics.[28,29] Clinical evidence suggests that statins impair the safety of colchicine,[30] as was again suggested by the first safety analyses of the COLCHICORT trial, showing that participants experiencing diarrhea were treated with statins twice as often as those who did not.[11] The explanation is expected to be that statins increase exposure to colchicine by its competitive metabolism, either through competition for CYP 3A4 (competitive metabolism clearance) or for *P*-gp (decreased biliary and enterocytic clearance) according to the type of statin.[29,31] This post hoc analysis further suggests, despite a lack of statistical power, that statins are associated with colchicine's gastrointestinal AEs. Around 20% of colchicine is estimated to be cleared through the kidneys, but evidence in the literature is contradictory on whether chronic kidney disease is associated with a higher risk of AEs with colchicine.[32,33] Our results indicate that even a mild decrease in eGFR could be associated with an increased risk of gastrointestinal AEs.

There seems to be a dual effect of conditions and cotreatments increasing colchicine exposure, resulting in improved efficacy but also an increased risk of gastrointestinal side effects. Prior studies have addressed the issue of factors associated with safety alone, but it was not known how they could also influence treatment response. With limitations inherent to the statistical power of the trial, this post hoc analysis suggests that the above-mentioned factors associated positively or negatively with colchicine safety are associated with its efficacy. For example, statins and older age to some extent seemed to be associated with a better treatment response but more gastrointestinal AEs, both potentially explained by increased exposure to colchicine. Diuretics, on the other hand, seemed to have a negative effect on colchicine's efficacy but were also associated with an increased risk of gastrointestinal AEs. These findings further question the ideal dose of colchicine according to age, renal function, and other current drugs at the time of the flare to provide the best possible balance between efficacy and safety. This concept is however in contradiction with the results of the AGREE trial performed in gout flares, which showed that high doses of colchicine (up to 4.8 mg/day) did not provide better rapid (24 hours) pain relief than lower doses (1.8 mg/day), despite an obvious higher exposure to colchicine.[19] Renal function and coprescriptions were however not reported in the AGREE trial, and participants were on average almost 40 years younger than those in the COLCHICORT trial, suggesting that these combined minor modulators of colchicine exposure may have a greater impact on older poly-medicated people.

The study has inherent limitations. First, sample size impacted the statistical power of the study and required a selection of variables of interest and limited the analysis of the drugs' individual efficacy, which was deliberately limited to an unadjusted univariable analysis often failing to reach statistical significance because of insufficient power, to avoid building invalid multivariable models. However, the robustness of the variables and outcomes prospectively collected in this first-ever trial in acute CPPD disease provided a strong offset for these drawbacks, allowing clinically relevant hypotheses to be drawn from these

observations, most of which were nonstatistically significant. Second, the prognostic value of the machine-learning generated classification needs to be considered with caution given the substantial number of misclassifications in crossvalidation. These results need to be validated in other cohorts or trials to confirm that they can be extrapolated in all contexts of acute CPP crystal arthritis.

This post hoc analysis of the COLCHICORT trial identified factors associated with a sustained response to colchicine and prednisone treatment for acute CPP crystal arthritis after a 2-day treatment regimen. Prednisone appeared to perform better in achieving pain relief in acute CPP crystal arthritis after two days of treatment, without requiring any further use of anti-inflammatory agents, and should be the first-line option for efficacy and safety reasons, except in case of type 2 diabetes. Renal function and comorbidities, inflammatory biomarkers, arthritis site, presence of underlying OA, and other current drugs are important determinants of treatment response and safety and should further guide the choice between them. Statins and age seem to have a dual effect on colchicine's efficacy and safety profiles, and further dose-finding studies adjusting on these factors are required.

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 Pascart 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. Pascart T, Filippou G, Lioté F, et al. Calcium pyrophosphate deposition disease. Lancet Rheumatol 2024;6(11):e791–e804.
2. Abhishek A, Neogi T, Choi H, et al. Review: unmet needs and the path forward in joint disease associated with calcium pyrophosphate crystal deposition. Arthritis Rheumatol 2018;70(8):1182–1191.
3. Rosenthal AK, Ryan LM. Calcium pyrophosphate deposition disease. N Engl J Med 2016;374(26):2575–2584.
4. Zhang W, Doherty M, Bardin T, et al. European League Against Rheumatism recommendations for calcium pyrophosphate deposition. Part I: terminology and diagnosis. Ann Rheum Dis 2011;70(4): 563–570.
5. Latourte A, Ea H-K, Richette P. Calcium pyrophosphate and basic calcium phosphate crystal arthritis: 2023 in review. Gout Urate Cryst Depos Dis 2024;2(2):101–107.
6. Sirotti S, Scanu A, Pascart T, et al. Calcium pyrophosphate crystal formation and deposition: Where do we stand and what does the future hold? Curr Rheumatol Rep 2024;26(10):354–365.
7. Tedeschi SK. A New Era for Calcium Pyrophosphate Deposition Disease Research: The First-Ever Calcium Pyrophosphate Deposition Disease Classification Criteria and Considerations for Measuring Outcomes in Calcium Pyrophosphate Deposition Disease. Gout Urate Cryst Depos Dis 2024;2(1):52–59.
8. McCarthy GM, Dunne A. Calcium crystal deposition diseases - beyond gout. Nat Rev Rheumatol 2018;14(10):592–602.
9. Pascart T, Latourte A, Tedeschi SK, et al. Features associated with different inflammatory phenotypes of calcium pyrophosphate deposition disease: A study using data from the International American College of Rheumatology/EULAR Calcium Pyrophosphate Deposition Classification Criteria Cohort. Arthritis Rheumatol 2024;76(12):1780–1788.
10. Abhishek A, Tedeschi SK, Pascart T, et al. The 2023 ACR/EULAR classification criteria for calcium pyrophosphate deposition disease. Arthritis Rheumatol 2023;75(10): 1703–1713.
11. Pascart T, Robinet P, Ottaviani S, et al. Evaluating the safety and short-term equivalence of colchicine versus prednisone in older patients with acute calcium pyrophosphate crystal arthritis (COLCHICORT): an open-label, multicentre, randomised trial. Lancet Rheumatol 2023;5(9):e523–e531.
12. Maravic M, Ea HK. Hospital burden of gout, pseudogout and other crystal arthropathies in France. Joint Bone Spine 2015;82(5): 326–329.
13. Ramonda R, Oliviero F, Galozzi P, et al. Molecular mechanisms of pain in crystal-induced arthritis. Best Pract Res Clin Rheumatol 2015; 29(1):98–110.
14. Zhang W, Doherty M, Pascual E, et al. EULAR recommendations for calcium pyrophosphate deposition. Part II: management. Ann Rheum Dis 2011;70(4):571–575.
15. Mandl P, D'Agostino MA, Navarro-Compán V, et al. 2023 EULAR recommendations on imaging in diagnosis and management of crystal-induced arthropathies in clinical practice. Ann Rheum Dis 2024;83(6):752–759.
16. Tedeschi SK, Becce F, Pascart T, et al. Imaging features of calcium pyrophosphate deposition disease: Consensus definitions from an international multidisciplinary working group. Arthritis Care Res (Hoboken) 2023;75(4):825–834.
17. Lydersen S. Statistical review: frequently given comments updated. Ann Rheum Dis 2025;84(5):660–663.
18. Neufeld AC, Gao LL, Witten DM. Tree-values: selective inference for regression trees. J Mach Learn Res 2022;23:23.
19. Terkeltaub RA, Furst DE, Bennett K, et al. High versus low dosing of oral colchicine for early acute gout flare: twenty-four-hour outcome of the first multicenter, randomized, double-blind, placebo-controlled, parallel-group, dose-comparison colchicine study. Arthritis Rheum 2010;62(4):1060–1068.
20. Leung YY, Haaland B, Huebner JL, et al. Colchicine lack of effectiveness in symptom and inflammation modification in knee osteoarthritis (COLKOA): a randomized controlled trial. Osteoarthritis Cartilage 2018;26(5):631–640.
21. Døssing A, Henriksen M, Ellegaard K, et al. Colchicine twice a day for hand osteoarthritis (COLOR): a double-blind, randomized, placebo-controlled trial. Lancet Rheumatol 2023;5(5):e254–e262.
22. Singh A, Molina-Garcia P, Hussain S, et al. Efficacy and safety of colchicine for the treatment of osteoarthritis: a systematic review and meta-analysis of intervention trials. Clin Rheumatol 2023;42(3): 889–902.
23. Kroon FPB, Kortekaas MC, Boonen A, et al. Results of a 6-week treatment with 10 mg prednisolone in patients with hand osteoarthritis (HOPE): a double-blind, randomised, placebo-controlled trial. Lancet 2019;394(10213):1993–2001.
24. Sharma L, Kwoh K, Lee JJ, et al. Development and validation of risk stratification trees for incident slow gait speed in persons at high risk for knee osteoarthritis. Ann Rheum Dis 2019;78(10):1412–1419.

25. Jatuworapruk K, Grainger R, Dalbeth N, et al. Development of a prediction model for inpatient gout flares in people with comorbid gout. Ann Rheum Dis 2020;79(3):418–423.
26. Rehberg M, Giegerich C, Praestgaard A, et al; MOBILITY, MONARCH, TARGET, and ASCERTAIN investigators. Identification of a rule to predict response to sarilumab in patients with rheumatoid arthritis using machine learning and clinical trial data. Rheumatol Ther 2021;8(4):1661–1675.
27. Bolognesi MM, Capitoli G, Galimberti S, et al. Dissecting the histological features of lupus nephritis highlights new common patterns of injury in class III/IV. Ann Rheum Dis 2022;81(12):1704–1711.
28. Pascart T, Richette P. Colchicine in gout: an update. Curr Pharm Des 2018;24(6):684–689.
29. Terkeltaub RA, Furst DE, Digiacinto JL, et al. Novel evidence-based colchicine dose-reduction algorithm to predict and prevent colchicine toxicity in the presence of cytochrome P450 3A4/P-glycoprotein inhibitors. Arthritis Rheum 2011;63(8):2226–2237.
30. Schwier NC, Cornelio CK, Boylan PM. A systematic review of the drug-drug interaction between statins and colchicine: patient characteristics, etiologies, and clinical management strategies. Pharmacotherapy 2022;42(4):320–333.
31. Kwon OC, Hong S, Ghang B, et al. Risk of colchicine-associated myopathy in gout: influence of concomitant use of statin. Am J Med 2017;130(5):583–587.
32. Bausson J, Keller N, Von Hunolstein JJ, et al. Safety and efficacy of colchicine in crystal-induced arthritis flare in 54 patients with severe chronic kidney disease. RMD Open 2024;10(1):e003872.
33. Bajpai R, Partington R, Muller S, et al. Prognostic factors for colchicine prophylaxis-related adverse events when initiating allopurinol for gout: retrospective cohort study. Rheumatology (Oxford). 2025;64(3):1147–1154.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 512–519
DOI 10.1002/acr.25649

American College of Rheumatology
Empowering Rheumatology Professionals

Clinical Outcomes of Patients Within the Rheumatoid Arthritis Care Pathway Cohort at a Tertiary Care Integrated Delivery Network: A Comparison to Usual Care

Tarun Sharma,[1] Tyson S. Barrett,[2] Teigan Dwyer,[2] Nancy Campbell,[1] Jessica Heintzinger,[1] Ellen Kraemer,[1] Adam Dore,[1] and Susan Manzi[1]

Objective. We aimed to compare clinical outcomes between patients in the Allegheny Health Network rheumatoid arthritis (RA) care pathway and patients receiving usual care.

Methods. The care pathway initiative implements guideline-based best practice alongside multidisciplinary team-based care. Clinical and insurance claims data were extracted to compare the proportion of patients in clinical disease activity index (CDAI)–based remission and evaluate health care utilization and per-member per-month (PMPM) costs of care.

Results. The RA care pathway cohort included 817 patients, and the usual care cohort included 3,651 patients. The care pathway cohort had a significantly higher proportion of patients achieving remission at 6, 12, and 24 months (33.5% vs 20.3%; hazard ratio 1.53, 95% confidence interval 1.11–2.10; P = 0.008 at 24 months), and on average, remission occurred 64 days earlier than usual care. This trend was also observed when including patients with low disease activity with those in remission and in a subgroup analysis of newly diagnosed RA. RA-related PMPM costs were significantly higher in the care pathway group and primarily related to higher baseline CDAI, comorbidities, and biologic/targeted synthetic disease modifying antirheumatic drug use and switching.

Conclusion. Patients in our RA care pathway were more likely to achieve remission than those in usual care. PMPM RA costs were higher compared to usual care. We plan to use a longer follow-up to investigate if improved clinical outcomes result in reduced direct and indirect costs, to study the impact of individual team-based care interventions, and to validate our findings in other populations.

INTRODUCTION

Rheumatoid arthritis (RA) is the most common form of systemic inflammatory arthritis and has a substantial societal effect in terms of costs, disability, and productivity.[1] Guidelines for managing RA recommend a treat-to-target approach, and compelling evidence supporting this method has made treat-to-target the main pharmacologic approach to managing RA.[2] However, implementing treat-to-target is fraught with several barriers.[3] Disease modifying antirheumatic drugs, either conventional synthetic or biologic and targeted synthetic (b/tsDMARDs), have revolutionized RA management, but achieving or maintaining remission can still be difficult due to individual patient factors. Social determinants of health (SDOH) can affect patient care and outcomes but are not directly addressed by the conventional treat-to-target pharmacologic approach. For example, factors such as smoking are implicated in both epigenetics and clinical outcomes in RA.[4,5] There is limited evidence on the impact of other SDOH or interventions addressing SDOH on outcomes for patients with RA.[3,6] The clinical impact of a RA care pathway that systematically addresses SDOH while implementing a guideline-based best practice approach has never been studied. We evaluated the efficacy of Allegheny Health Network's (AHN) RA care pathway by measuring clinical outcomes and health care utilization costs.

A care pathway must be multidisciplinary, employ guidelines and evidence, provide treatment as a plan/pathway/algorithm, and standardize care for a specific population.[7] Keeping in mind these care pathway principles and the known care gaps affecting

[1]Tarun Sharma, MD, Nancy Campbell, RN, BSN, Jessica Heintzinger, Ellen Kraemer, MBA, Adam Dore, DO, Susan Manzi, MD, MPH, Allegheny Health Network, Pittsburgh, Pennsylvania; [2]Tyson S. Barrett, PhD, Teigan Dwyer, MPH, Highmark Health, 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.25649).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25649.

Address correspondence via email to Tarun Sharma, MD, at tarun.sharma@ahn.org.

Submitted for publication January 27, 2025; accepted in revised form September 2, 2025.

SIGNIFICANCE & INNOVATIONS

- Rheumatoid arthritis (RA) care pathways implementing guideline-based RA best practices alongside team-based care could lead to more patients achieving remission and faster time to remission compared to usual care.
- Increased biologic and targeted synthetic disease modifying antirheumatic drug use and switching could be a factor driving costs higher, especially if patients start with higher disease activity and comorbidities.

patients with RA, the RA care pathway initiative was launched at AHN, a tertiary care academic medical center, in collaboration with the integrated payor Highmark Health. We describe the design and implementation of the AHN RA care pathway and compare the clinical outcomes, health care utilization, and costs between patients in the care pathway and patients receiving usual care.

PATIENTS AND METHODS

Patients. The care pathway was implemented at two clinical sites, including one urban and one suburban site, with a total of 13 providers (9 rheumatologists and 4 physician extenders). A comparison cohort included patients in the same health system who were diagnosed with RA but receiving usual care at four clinical sites, including three suburban and one urban site, with a total of six rheumatologists and four extenders.

Implementation phases are outlined in Figure 1. RA care transformation efforts started in 2020 quarter (Q)1 by an expert panel attempting to answer the fundamental question: what does ideal RA care look like from a patient perspective, and how do we strive to achieve it with every patient encounter? The first phase of implementation was the Definition Phase (2020 Q1) and involved defining outcomes, such as RA classification criteria, clinical outcome measures, and guideline-based RA care best practices. This required a detailed needs assessment, process mapping of current and desired workflows, and identification of barriers and facilitators with the help of key stakeholders and institutional champions. The second stage was the Design and Collect Phase (2020 Q2–2021 Q4) and involved process design and implementation of team-based care and prospective data capture tools at point of care with the help of our implementation team and data experts. This involved weekly team meetings and implementing small-scale changes in a cyclical manner at our first site, the Autoimmunity Institute. Rheumatology providers or the nurse navigators were required to validate the RA diagnosis using American

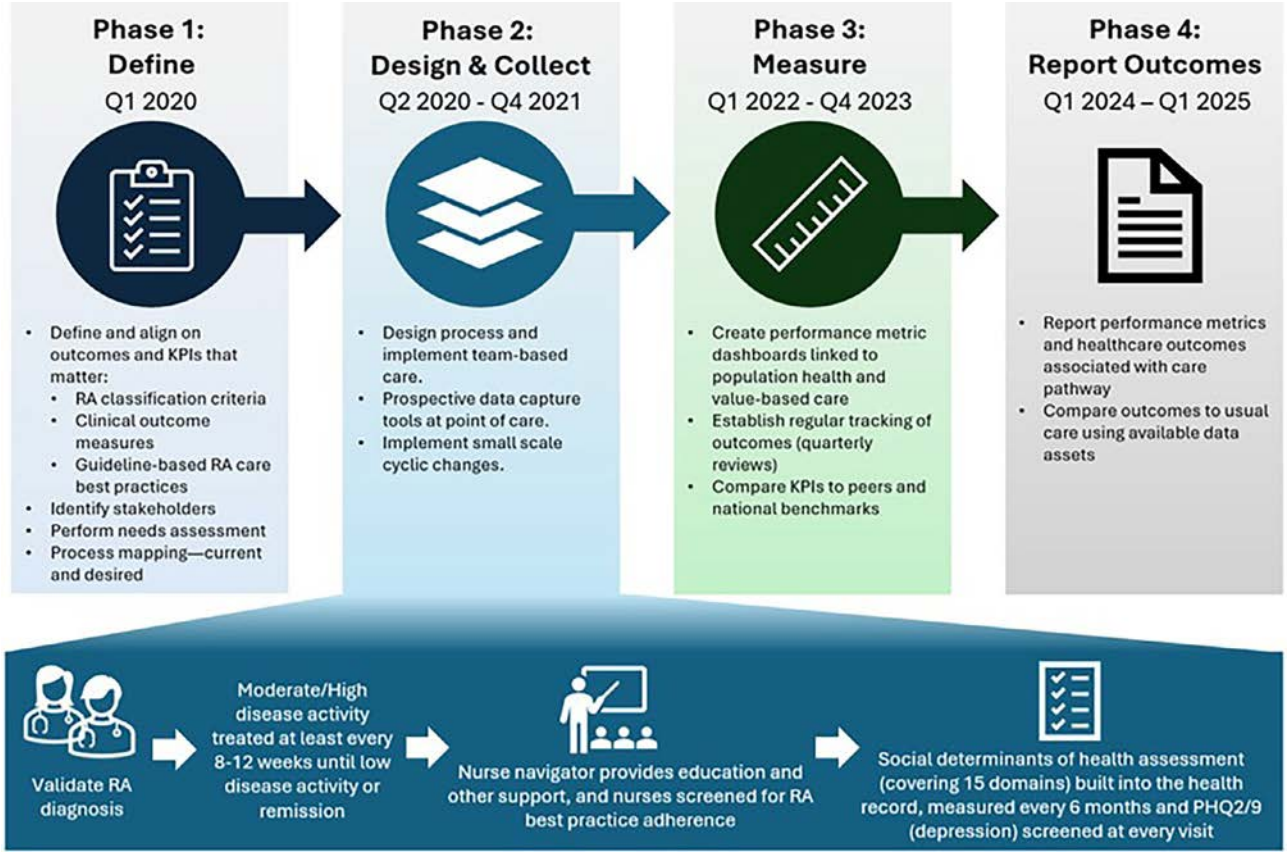

Figure 1. RA care pathway implementation details. KPI; key performance indicator; PHQ, Patient Health Questionnaire; Q, quarter; RA, rheumatoid arthritis.

College of Rheumatology (ACR)/EULAR 2010 RA classification criteria and added the patient to the care pathway. Providers followed the ACR 2021 guidelines and best practices for managing RA, including the treat-to-target approach, and preventive and safety-related best practices outlined in Table 1. Patients with moderate-to-high disease activity were observed at least every 8 to 12 weeks until remission or low disease activity was achieved. The nurse navigator facilitated RA-related education and addressed personalized needs for every patient added to the pathway. This included telephonic follow-ups after two weeks and at least every three months after addition to the care pathway, being identified as having moderate-to-high clinical disease activity index (CDAI) activity or initiation/modification of b/tsDMARD. At follow-up, nurses screened for RA best practice adherence, and providers were notified of any best practice care gaps.

Team-based care focused on patient-centric aspects of RA care, such as productive interactions between an informed, activated patient (via patient education on RA disease and therapeutics), and a prepared, proactive care pathway team that consisted of a rheumatology provider, nurse navigator, clinical pharmacist, behavioral health specialist, registered dietician, and social worker. The team identified and addressed SDOH through a questionnaire that was integrated into the electronic health record (EHR) and administered to patients (Supplementary Table S1). The questionnaire covered the following 15 domains: alcohol and drug use, tobacco use, access and affordability, employment, financial strain, health literacy, depression, food insecurity,

Table 1. Comparison of demographics and performance metrics of RA care pathway cohort vs usual care*

Characteristic	Care pathway		Usual care		*P* value
	Total	Value	Total	Value	
Baseline demographics					
Gender, female patients	817	630 (77)	3,651	2,767 (76)	0.4
Age	817	62 (53–71)	3,651	65 (55–72)	<0.001
White	817	647 (79)	2,517	2,327 (92)	<0.001
Education: college or higher (Highmark patients only)	369	98 (27)	937	245 (24)	0.3
Newly diagnosed RA	817	127 (16)	1,025	129 (13)	0.2
Smoker, current	817	107 (13.0)	3,651	318 (8.7)	<0.001
Seropositivity (RF and/or ACPA)	409	257 (63)	1,645	847 (51)	<0.001
Comorbidities					
CHF	817	81 (9.9)	3,651	286 (7.8)	0.050
PVD	817	159 (19)	3,651	548 (15)	0.002
COPD	817	129 (16)	3,651	416 (11)	<0.001
PUD	817	42 (5.1)	3,651	108 (3.0)	0.002
Diabetes	817	182 (22)	3,651	514 (14)	<0.001
Cancer	817	175 (21)	3,651	619 (17)	0.003
Depression	817	269 (33)	3,651	736 (20)	<0.001
Disease management best practice adherence					
RA classification criteria complete	817	817 (100)	-	-	-
CDAI completion: baseline (first 3 mo of study or immediately before care pathway)	817	477 (58)	3,651	900 (25)	<0.001
Baseline CDAI value	477	9 (3–18)	900	6 (2–11)	<0.001
Baseline CDAI remission	477	96 (20)	900	244 (27)	0.004
Number of CDAI measures	817	3 (1–5)	3,651	0 (0–1)	<0.001
CDAI completion: final (final 3 mo of study)	817	549 (67)	3,651	830 (23)	<0.001
b/tsDMARDs use (Highmark patients only)	364	171 (47)	1,001	360 (36)	<0.001
b/tsDMARD ≥1 switch (Highmark patients only)	364	130 (36)	1,001	238 (24)	0.018
Prevention and safety best practice adherence					
Team-based care referrals	817	117 (14)	3,651	130 (3.6)	<0.001
PHQ-2/9 depression screening at least once	817	776 (95)	3,651	2,346 (64)	<0.001
PHQ-9 score, mean ± SD	812	1.03 ± 3.22	2,345	0.94 ± 3.35	0.030
SDOH ≥4 domains available	817	548 (67)	3,651	1,804 (49)	<0.001
Influenza vaccination annually	817	537 (66)	3,651	2,222 (61)	0.010
Pneumococcal vaccination within 5 y	134	54 (40)	3,324	1,369 (41)	0.8
COVID-19 vaccination	817	646 (79)	3,651	2,857 (78)	0.6
Tuberculosis testing prebiologic within 1 y	161	88 (55)	1,648	1,174 (71)	<0.001
Hepatitis B testing prebiologic within 5 y	161	103 (64)	1,648	967 (59)	0.2
Lipid panel annually	817	242 (30)	3,651	595 (16)	<0.001

* Statistics presented as n (%) or median (interquartile range) unless otherwise indicated. Where not listed for usual care, data are not available. Blended sample includes clinical, claims, utilization, and cost data. ACPA, anti–citrullinated protein antibody; b/tsDMARD, biologic and targeted synthetic disease modifying antirheumatic drug; CDAI, clinical disease activity index; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; PHQ, Patient Health Questionnaire; PUD, peptic ulcer disease; PVD, peripheral vascular disease; RA, rheumatoid arthritis; RF, rheumatoid factor; SDOH, social determinants of health.

transportation, housing stability, stress, physical activity, safety, social connections, and veteran status. A subset of these domains was previously studied at our health system and found to correspond well with the social vulnerability index.[8] The nurse navigator or provider requested a referral to the specific care path team member based on the SDOH needs of the individual patient. For example, a clinical pharmacist provided medication management and monitoring and smoking cessation services; a behavioral health counselor provided short-term counseling, patient/family support, and coping strategies; a social worker addressed financial concerns, care affordability, support to resolve food, housing, and transportation needs; and a registered dietician helped define dietary goals and provided RA-specific dietary education. Our SDOH prevalence study showed that patients with moderate-to-severe RA disease activity had higher rates of depression,[5] and another preliminary baseline analysis showed that team-based care interventions addressing SDOH were associated with improved median CDAI scores by three points after two touchpoints. Therefore, Patient Health Questionnaire (PHQ)-29 depression screening was performed at every visit, whereas other domains were measured at least every six months. After successfully piloting these efforts at the first site, the RA pathway initiative was scaled up to the second clinical site. Ongoing training and education were provided.

The third and fourth phases were measurement and reporting, respectively. The Measurement Phase (2022 Q1–2023 Q4) used performance metric dashboards linked to population health and value-based care. Based on our previous study findings that longitudinally tracking RA performance metrics and sharing data transparently was vital to implementing change, we shared quarterly dashboards for patients seen between January 1, 2022, and December 31, 2023, at staff meetings with predefined key performance indicators, including comparison to peers and national registry benchmarks where available.[9,10] These are detailed in Table 1. The Reporting Phase (2024 Q1–2025 Q1) involved reporting on the comparison of clinical and other outcomes of interest between the RA care pathway and usual care.

Data sources. Clinical information was extracted from the EHR and cross-referenced with insurance claims information for patients continuously covered by Highmark Health plans. Data were linked internally as both the payor and provider are part of an integrated health system. Gender and race (as self-identified by patients) were also extracted from the EHR. Providers specifically designated patients with RA who were entered into the RA care pathway as either newly diagnosed or established at point of care. Patients in the usual care group were identified retrospectively from our AHN RA EHR registry and designated as newly diagnosed with RA if they had no RA-related claim for a period of at least 24 months going into their first visit during study phase 2 or 3. They were designated as established RA if they had at least two rheumatology visits using *International Classification of Diseases-10* M05.xx or M06.xx going into the study phase 2 or 3.

Outcome measures. The primary outcome was the proportion of patients in CDAI-based remission (ie, CDAI score ≤2.8) at 3, 6, 12, and 24 months. Secondary outcomes included time to remission, health care utilization (emergency room and inpatient), and per-member per-month (PMPM) costs of care (RA-related and medication-related). We compared clinical outcomes, health care utilization, and PMPM costs of care of the RA care pathway cohort with usual care.

Statistical analyses. Patient characteristics were summarized as mean ± SD and median (interquartile range [IQR]) for continuous variables and frequency (%) for categorical variables. A standard *t*-test was used for continuous, normally distributed variables, and the nonparametric Wilcoxon rank sum (Mann-Whitney *U*) test was used for continuous variables that did not meet normality requirements. A χ^2 or Fisher's exact test was used for categorical variables. For analyses regarding remission, three analyses were used. First, a snapshot of remission among patients still enrolled at three months, six months, one year, and two years analyzed events up to that date among all those still enrolled at that point in time. Second, linear regression analyses of patients who reached remission predicted the number of days to remission between groups, controlling for age, gender, race, and baseline CDAI scores. Third, Kaplan-Meier curves and Cox proportional hazard models estimated time to remission for individuals not already in remission at baseline and had at least two CDAI scores. The time to event analyses were calculated for all patients and a subgroup of patients with newly diagnosed RA. Where appropriate, assumptions of the statistical tests were checked for violations.

Clinical outcomes were compared using the total study population, whereas cost and utilization outcomes were compared within the Highmark-enrolled subgroups. Institutional review board approval was obtained, and informed consent was waived. The statistical software programs used were R version 4.2.3 and SAS version 9.4. A *P* value ≤0.05 was considered statistically significant.

RESULTS

The RA care pathway cohort included 817 patients at end of the measurement period, of which 364 had continuous Highmark enrollment data. The usual care cohort included 3,651 patients from the RA registry, of which 1,001 patients had Highmark enrollment data. Details of the cohorts can be found in the CONSORT diagram (Figure 2). Some data sources had missing values for those in the sample, and therefore all sample sizes are reported.

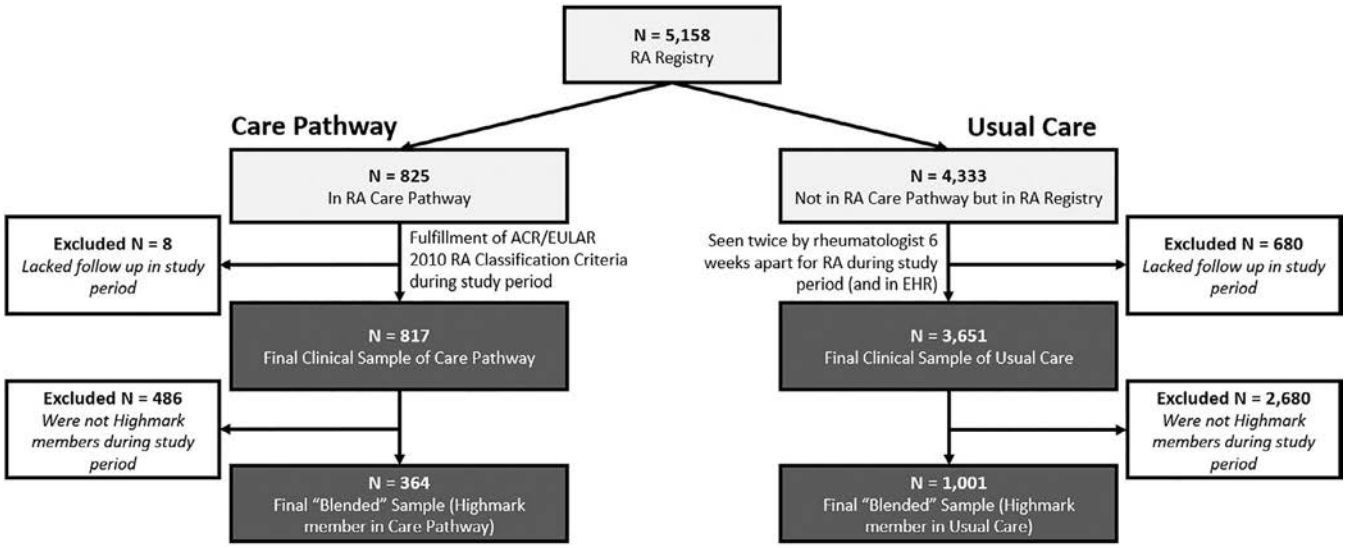

Figure 2. Consolidated Standards of Reporting Trials diagram showing care pathway and usual care cohorts for patients with RA. Blended sample includes clinical, claims, utilization, and cost data. ACR, American College of Rheumatology; EHR, electronic health record; RA, rheumatoid arthritis.

Patient demographics differed across several metrics at baseline (Table 1). There were several differences in comorbidities between the two cohorts, such as higher rates of congestive heart failure, diabetes, and depression in the care pathway cohort. Those in the care pathway tended to be seropositive. Baseline CDAI values were higher in the care pathway cohort, and care pathway patients had more CDAI measures than the usual care cohort. A higher proportion of care pathway patients received multidisciplinary team-based care referrals compared to the usual care group. Of patients using b/tsDMARDs, a significantly higher number of care pathway patients had at least one b/tsDMARD switch compared to usual care patients (36% [130 of 364] vs 24% [238 of 1,001]; $P = 0.018$). In addition, there was no difference in the proportion of days covered (adherence measure) between the groups ($P = 0.585$), although the care pathway group tended to have used b/tsDMARDs for longer (median 900 days vs 706 days; $P < 0.001$).

Remission rates were different between the care pathway and usual care groups (Table 2). A significantly higher proportion of patients reached remission in the care pathway cohort compared to the usual care cohort at one-year (20.8% vs 12.4%; $P = 0.001$) and two-year (33.5% vs 20.3%; $P = 0.008$) follow-up. Although not statistically significant ($P = 0.081$), regression analyses showed that patients in the care pathway group reached remission on average 64 days faster (after controlling for age, gender, race, and baseline CDAI scores). Similarly, on inclusion of low disease activity (CDAI score ≤10) as an acceptable target, a significantly higher proportion of patients achieved low disease activity or remission in the care pathway cohort compared to the usual care cohort at one-year (48.8% vs 18.1%; $P < 0.001$) and two-year (68.8% vs 16.7%; $P < 0.001$) follow-up.

In a subgroup analysis of only patients newly diagnosed with RA, 35.4% were in remission within one year, and 45.5% were in remission by two years in the care pathway group. This was more frequent than in patients with newly diagnosed RA in the usual care group; however, this was not statistically significant (24.6% by 1 year, $P = 0.662$ and 40% by end of year 2, $P = 0.434$). Additionally, although not significant ($P = 0.135$), these patients reached remission approximately 124 days faster than those in usual care (after controlling for age, gender, race, and baseline CDAI scores). When including low disease activity with remission outcomes, a significantly higher proportion of newly diagnosed patients achieved low disease activity or remission in the care pathway cohort compared to the usual care cohort at one-year (67.7% vs 31.1%; $P < 0.001$) and two-year (86.3% vs 22.5% $P < 0.001$) follow-up.

The usual care group had more patients in remission at the start of the study. To properly assess and compare time to remission between groups, we excluded patients already in remission at the start of the study period (Kaplan-Meier survival curve; Figure 3). This showed a significant difference in time to remission between the groups ($P = 0.005$). The steepest slope was new patients with RA in the care pathway, followed by new patients with RA in the usual care cohort.

PMPM costs were higher in the care pathway for overall RA costs ($P < 0.001$), with median per-month costs for care pathway at $48.90 (IQR, 23.9–634.0) and usual care at $25.70 (IQR, 11.3–62.5) for the first 12 months. For those on b/tsDMARDs, the per-month costs in the first 12 months were marginally higher ($P = 0.060$) for the care pathway (median, $3,619; IQR, (3–4,867) than the usual care (median, $2,550; IQR, 3–4,169) patients. Again, for those using b/tsDMARDs in months 13 to 24, the care

Table 2. Comparison of clinical outcomes of the care pathway cohort vs usual care*

Variable	Pathway, n (%)	Usual care, n (%)	Unadjusted *P* value	Adjusted effect size, HR (95% CI)	Adjusted *P* value
New and established RA					
Remission					
3 mo	17 (3.1)	14 (2.0)	0.266	1.01 (0.75–1.35)	0.925
6 mo	45 (8.4)	36 (5.6)	0.076	1.29 (0.98–1.71)	0.066
12 mo	97 (20.8)	50 (12.4)	0.001	1.73 (1.33–2.27)	<0.001
24 mo	88 (33.5)	61 (20.3)	<0.001	1.53 (1.11–2.10)	0.008
Low disease activity or remission					
3 mo	54 (9.9)	36 (5.1)	<0.001	0.95 (0.81–1.12)	0.571
6 mo	147 (27.6)	78 (12.2)	<0.001	0.96 (0.94–1.29)	0.233
12 mo	228 (48.8)	73 (18.1)	<0.001	1.23 (1.03–1.46)	0.023
24 mo	181 (68.8)	50 (16.7)	<0.001	1.35 (1.10–1.68)	0.005
New RA only					
Remission					
3 mo	4 (4.7)	5 (5.9)	>0.99	1.74 (0.56–5.38)	0.332
6 mo	11 (13.8)	10 (12.8)	>0.99	1.02 (0.35–2.94)	0.969
12 mo	23 (35.4)	15 (24.6)	0.261	0.82 (0.32–2.03)	0.662
24 mo	10 (45.5)	16 (40.0)	0.883	1.62 (0.48–5.43)	0.434
Low disease activity or remission					
3 mo	12 (14.1)	11 (12.9)	>0.99	3.06 (1.36–6.88)	0.006
6 mo	36 (45.0)	20 (25.6)	0.017	2.24 (1.26–3.97)	0.006
12 mo	44 (67.7)	19 (31.1)	<0.001	2.95 (1.70–5.12)	<0.001
24 mo	19 (86.3)	9 (22.5)	<0.001	2.53 (1.27–5.05)	0.008

* Adjusted models included age, gender, race, and baseline clinical disease activity index (CDAI) scores controls. Remission excludes patients already in remission and remission/low disease activity excludes those in remission/low disease activity before the start date and required at least 2 CDAI measures. CI, confidence interval; HR, hazard ratio; RA, rheumatoid arthritis.

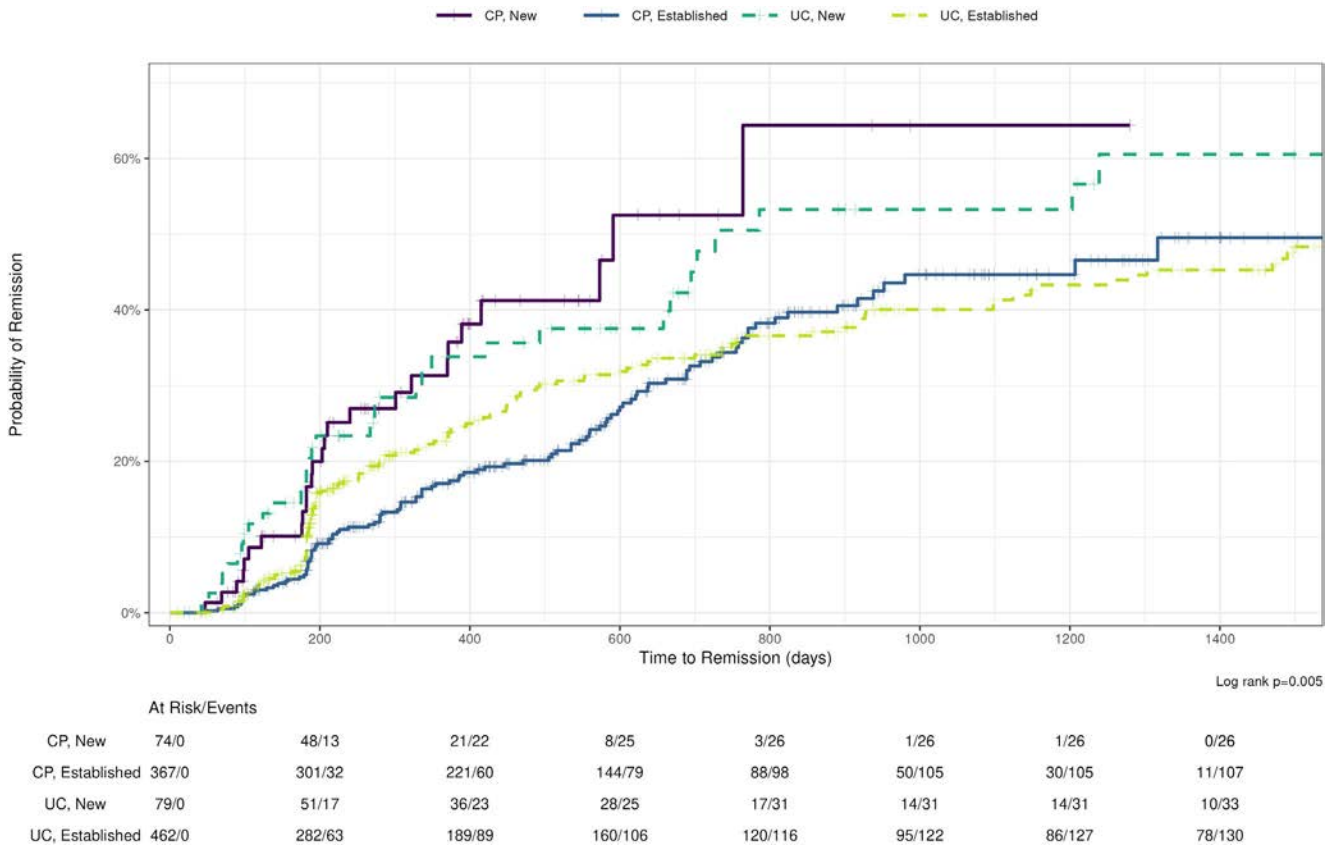

Figure 3. Kaplan-Meier curves of time to remission in patients in CP vs UC by new or established rheumatoid arthritis diagnosis, excluding individuals already in remission before or on the start date. CP, care pathway cohort; UC, usual care. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25649/abstract.

pathway had higher ($P = 0.038$) costs (median, $4,738; IQR, 529–6,351) than usual care (median, $2,464; IQR, 2–4,509). A similar pattern existed for RA costs in months 13 to 24, with $55.70 (IQR, 20.1–333.8) for the care pathway and $28.80 (IQR, 12.2–91.4) for usual care ($P < 0.001$).

No significant differences existed for inpatient admissions and RA-related or infection-related emergency department (ED) visits for the first 12 months, particularly after adjusting for age, gender, race, and baseline CDAI scores. These remained nonsignificant for months 13 to 24, with drops in utilization, particularly for pathway patients (from 9.6% to 5.9%), in both unadjusted and adjusted comparisons. The only significant difference in the first 12 months was for general ED visits (31.7% for pathway, 20.7% for usual care; $P < 0.001$, adjusted $P = 0.001$). This difference decreased to nonsignificant levels by months 13 to 24 (28.1% for pathway, 22.8% for usual care; $P = 0.163$, adjusted $P = 0.117$), with care pathway patients less likely and usual care patients more likely to visit the ED compared to the first 12 months.

Health care costs and utilization were proportionate to RA disease activity in both groups. For the first 12 months, patients in low disease activity/remission had fewer ED visits (30% vs 21%; $P = 0.008$), fewer PMPM health care encounters (3.6 vs 3.1; $P = 0.019$), and fewer PMPM prescription fills (21 vs 18; $P = 0.026$). For 13 to 24 months, there were no significant differences across these comparisons (likely in part due to small sample sizes).

DISCUSSION

In our study, patients in our RA care pathway had significantly better clinical outcomes than patients receiving usual care. There was a significantly higher proportion of patients in remission at 6, 12, and 24 months in the care pathway group than the usual care group. Additionally, the average time to remission was shorter in care pathway patients than usual care patients, both before and after adjustment for age, gender, race, and baseline CDAI scores. A subgroup analysis of patients with newly diagnosed RA yielded similar trends in proportion of patients in remission and time to remission. The care pathway cohort had more patients with seropositive RA, higher comorbidities, and higher baseline CDAI scores than the usual care group. PMPM costs of the care pathway were higher when compared to usual care, which could be due to several factors, such as higher baseline CDAI, higher comorbidity burden, higher b/tsDMARDs use, and b/tsDMARDs switching. ED visits were higher in the care pathway group during the first 12 months but were not related to RA or infection and reduced to nonsignificant levels during months 13 to 24. Our data suggest that our care pathway program is successful in increasing the proportion of patients in remission and reducing time to remission compared to usual RA care.

To our knowledge, there is no other study on implementing a RA care pathway with a SDOH-focused, multidisciplinary team alongside guideline-based best practice. In 2022, Gukova et al[11] proposed an interdisciplinary Calgary early RA care pathway that included referral triage, diagnosis, and management; however, this care pathway was primarily focused on patients with early RA, and the implementation and outcomes data are not yet reported. Nurse-led RA programs have demonstrated encouraging results in the literature, but these programs only involved nurses assuming their own patient caseloads and providing direct patient services rather than a team-based approach.[12] Our RA care pathways were not primarily nurse-led, but the nurses assumed patient care related to RA as part of a team and also addressed SDOH and care gaps between visits.

Our study has several strengths. A complex care pathway that includes best practice implementation and a multidisciplinary, team-based care approach is best studied within an integrated delivery network like AHN to better assess the long-term clinical and financial outcomes. Our care pathway consisted of a well-defined and sizable population of almost 900 patients who met RA classification criteria. These patients were compared to a usual care cohort of around 3,000 patients, and all patients had claims, pharmacy, and dispense data, along with cost and utilization outcomes. The patient-centric, phased approach of defining outcomes, multidisciplinary team-building and design implementation, data collection, and dashboard development make the AHN RA care pathway a robust and well-designed program. Our subgroup analysis of a newly diagnosed RA cohort with improved clinical outcomes also further reaffirms our results.

Our study has limitations. First, findings from our resource-heavy care pathway initiative may not be generalizable to the community practice setting because community practices may not have the same level of access to multidisciplinary care teams, EHR integration, and data analytic resources. Our PMPM calculations did not include costs of team-based care implementation, which was a part of a value-based chronic care model at our integrated health system. Second, inherent selection bias could have been introduced when comparing different clinical sites' patient populations and performance metrics of two different groups of providers albeit within the same practice. Third, care pathway implementation came with higher costs of care, which could be attributed to a sicker RA cohort with higher baseline CDAI, higher comorbidity burden, and higher b/tsDMARD use with more switches between b/tsDMARD therapies further escalating costs. Finally, patients with RA were defined as newly diagnosed or established using either insurance claims or clinical EHR data, leading to differences in how the care groups were classified. For example, newly diagnosed or established patients were identified as such at point of care after fulfillment of the RA classification criteria in the care pathway group in an effort to ensure we enroll patients with a validated diagnosis into the pathway due to its resource intensive structure with the various clinical and data

teams. However, in the usual care group, these patients were identified and classified retrospectively from within the RA registry. Even though we tried to mitigate this by using previously validated algorithms with proven accuracy for capturing incident or established RA, there could be inherent bias due to this discrepancy between the two groups.

Our study shows that our RA care pathway program was successful in improving RA patient outcomes, such as proportion of patients in clinical remission and time to remission compared to usual care. Future efforts to improve RA outcomes will include a focus on the subset of patients with moderate-high disease activity and study the relative impact of different team-based care interventions within the care pathway cohort. A longer follow-up is planned to investigate if the improved clinical outcomes result in reduced direct and indirect costs associated with improved health and quality of life. If our findings are confirmed in other health care settings, implementing such an RA care pathway model could improve the overall health and quality of life of individuals living with RA.

ACKNOWLEDGMENTS

The authors thank Sarah Carey, MS, Jade Chang, BA, and Jacalyn Newman, PhD, of Allegheny Health Network's Health System Publication Support Office (HSPSO) for their assistance in editing and formatting the manuscript. The HSPSO is funded by Highmark Health, and all work was done in accordance with Good Publication Practice guidelines (http://www.ismpp.org/gpp3).

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 Sharma 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. Lee DM, Weinblatt ME. Rheumatoid arthritis. Lancet 2001;358(9285):903–911.
2. Fraenkel L, Bathon JM, England BR, et al. 2021 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2021;73(7):1108–1123.
3. Gossec L, Bessette L, Xavier RM, et al. Barriers to, facilitators of, and interventions to support treat-to-target implementation in rheumatoid arthritis: a systematic review. Arthritis Care Res (Hoboken) 2024;76(12):1626–1636.
4. Masdottir B, Jónsson T, Manfredsdottir V, et al. Smoking, rheumatoid factor isotypes and severity of rheumatoid arthritis. Rheumatology (Oxford) 2000;39(11):1202–1205.
5. Wiemer N, Yin Y, Sharma T. Social determinants of health and clinical outcomes in rheumatoid arthritis: a single center study [abstract]. Arthritis Rheumatol 2023;75(suppl 9). https://acrabstracts.org/abstract/social-determinants-of-health-and-clinical-outcomes-in-rheumatoid-arthritis-a-single-center-study/.
6. Stuifbergen AK, Morris M, Jung JH, Pierini D, Morgan S. Benefits of wellness interventions for persons with chronic and disabling conditions: a review of the evidence. Disabil Health J 2010;3(3):133–145.
7. Lawal AK, Rotter T, Kinsman L, et al. What is a clinical pathway? Refinement of an operational definition to identify clinical pathway studies for a Cochrane systematic review. BMC Med 2016;14(1):35.
8. Brignone E, LeJeune K, Mihalko AE, et al. Self-reported social determinants of health and area-level social vulnerability. JAMA Netw Open 2024;7(5):e2412109.
9. Roberts ET, Schmajuk G, Li J, et al. Latent tuberculosis screening among new users of a biologic or targeted synthetic disease-modifying antirheumatic drug: gaps in screening overall and among Janus kinase inhibitors. Arthritis Care Res (Hoboken) 2024;76(7):1037–1044.
10. Sharma T, Wetherington K, Mcaninch L, et al. Factors associated with clinical disease activity index adoption at an academic rheumatology practice [abstract]. Arthritis Rheumatol 2019;71(suppl 10). https://acrabstracts.org/abstract/factors-associated-with-clinical-disease-activity-index-adoption-at-an-academic-rheumatology-practice/.
11. Gukova X, Hazlewood GS, Arbillaga H, et al. Development of an interdisciplinary early rheumatoid arthritis care pathway. BMC Rheumatol 2022;6(1):35.
12. de Thurah A, Esbensen BA, Roelsgaard IK, et al. Efficacy of embedded nurse-led versus conventional physician-led follow-up in rheumatoid arthritis: a systematic review and meta-analysis. RMD Open 2017;3(2):e000481.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 520–527
DOI 10.1002/acr.25652

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Serum Soluble Mediator Signatures of Lupus Nephritis: Histologic Features and Response to Treatment

Andrea Fava,[1] Catriona A. Wagner,[2] Carla J. Guthridge,[2] Susan Macwana,[2] Wade DeJager,[2] Melissa E. Munroe,[2] Peter Izmirly,[3] H. Michael Belmont,[3] Betty Diamond,[4] Anne Davidson,[4] Paul J. Utz,[5] Michael H. Weisman,[6] Philip M. Carlucci,[3] Maria Dall'Era,[7] Kenneth Kalunian,[8] Chaim Putterman,[9,10] Jennifer Anolik,[11] Jennifer L. Barnas,[11] David Wofsy,[7] Diane Kamen,[12] Richard A. Furie,[13] Deepak A. Rao,[14] the Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus Network, Michelle Petri,[1] Joel M. Guthridge,[2] Jill Buyon,[3] and Judith A. James[2,15]

Objective. Lupus nephritis (LN) management remains challenging, and novel noninvasive biomarkers are needed. This study quantified serum soluble mediators in the Accelerating Medicines Partnership (AMP) LN cohort to identify biomarkers of histologic features and treatment response.

Methods. Patients with systemic lupus erythematosus (SLE) (n = 268) undergoing clinically indicated kidney biopsies (urine protein/creatinine ratio [UPCR] ≥ 0.5) were recruited through the AMP Rheumatoid Arthritis and SLE Network. Serum was collected at biopsy and 3-, 6-, and 12-month postbiopsy, alongside samples from 22 healthy controls. Concentrations of 66 immune mediators were quantified using xMAP multiplex assays, and (TACE) measured by enzyme-linked immunosorbent assay. Seven mediators with >95% values below detection limits were excluded from analyses. Bootstrapped least absolute shrinkage and selection operator (LASSO) regression identified proliferative LN (class III/IV ± V) predictors from baseline mediators. Associations with 12-month treatment response (complete/partial vs no response) were tested using three-month changes in LASSO-selected mediators and UPCR via logistic regression. Molecular clustering of mediator profiles was performed to identify LN subgroups.

Results. Proliferative patients with LN (class [III or IV] ± V; n = 160) displayed a distinct mediator profile compared with nonproliferative LN (class I, II, or V; n = 96). LASSO regression identified 20 mediators predictive of proliferative LN (areas under the curve, 0.82; 95% confidence interval [CI], 0.81–0.91), including elevated syndecan-1, tumor necrosis factor receptor type I, tumor necrosis factor receptor type II, and vascular cell adhesion molecule 1 (VCAM-1), as well as decreased CCL3//macrophage inflammatory protein 1α, CD40 ligand, and interleukin-5 levels. Among proliferative patients with LN, 3-month reductions in syndecan-1 and VCAM-1, mediators associated with intrarenal LN activity and/or chronicity, predicted the 12-month treatment response. A model incorporated these reductions and a decline in UPCR-predicted treatment response in proliferative LN (0.90; 95% CI, 0.82–0.98). Molecular clustering revealed four distinct LN subgroups with unique soluble mediator signatures and clinical features not captured by histology alone.

Conclusion. Serum soluble mediators, particularly syndecan-1 and VCAM-1, reflect LN histologic activity, and early decreases predict treatment response, supporting their potential use as noninvasive longitudinal biomarkers. The substantial heterogeneity within LN highlights the potential for biomarker-guided reclassification to advance precision medicine approaches.

INTRODUCTION

Lupus nephritis (LN) is a serious complication of systemic lupus erythematosus (SLE), causing substantial morbidity and early death, with increased prevalence and poorer outcomes in patients of non-European ancestry. Currently, LN is classified histologically based on inflammation and tissue injury using the International Society of Nephrology/Renal Pathology Society (ISN/RPS) criteria,[1] as well as the National Institutes of Health (NIH) activity indices (NIH-AI) and NIH chronicity indices

This work was supported by the Accelerating Medicines Partnership in rheumatoid arthritis and systemic lupus erythematosus network. Accelerating Medicines Partnership in rheumatoid arthritis and systemic lupus erythematosus is a public–private partnership (AbbVie Inc, Arthritis

SIGNIFICANCE & INNOVATIONS

- A panel of 20 serum immune mediators can distinguish between proliferative and nonproliferative lupus nephritis (LN).
- Early reductions in syndecan-1 and vascular cell adhesion molecule 1 are predictive of 12-month treatment response in proliferative patients with LN.
- Serum soluble mediators may serve as noninvasive, longitudinal biomarkers of LN histologic features and treatment response.

(NIH-CI).[2] However, histologic class and activity can change over time,[3] and kidney biopsies pose practical challenges for longitudinal monitoring. These classification systems aid in the selection of treatment decisions. However, despite recent advances, treatment responses are suboptimal and often vary within histologic classes.[4] Therefore, noninvasive markers of renal inflammation, tissue damage, and treatment response are needed to augment and improve treatment decisions and clinical care.

Many soluble mediators, including cytokines and chemokines, are elevated in the blood of patients with LN, suggesting that serological proteins may represent noninvasive biomarkers for LN classification and treatment personalization. Although some studies have demonstrated an association of several serological soluble mediators with the NIH-AI and NIH-CI,[5–7] studies on the association of mediators with histologic class or treatment response are limited and have focused on a small set of markers, which may inadequately capture disease heterogeneity. Furthermore, studies focused on racial and ethnic groups disproportionately affected by LN are less frequent. Therefore, in this study, we investigated the levels of 67 serum soluble mediators in a longitudinal, racially diverse cohort of patients with LN to determine early, noninvasive biomarkers of histologic class and one-year treatment response.

MATERIALS AND METHODS

Study population. Patients with SLE undergoing a clinically indicated kidney biopsy (urine protein/creatinine ratio [UPCR] ≥ 0.5) were recruited as part of the Accelerating Medicines Partnership rheumatoid arthritis and SLE network, as previously described (Appendix A).[8] Kidney biopsies were evaluated according to ISN/RPS classification[1] and the NIH-AI and NIH-CI[2] by a central renal pathologist. Healthy controls were recruited through the Oklahoma Medical Research Foundation's Oklahoma Rheumatic Disease Research Cores Center and New York University Medical Center. Demographic and clinical characteristics of the study population were previously described.[8]

All participants provided written informed consent before study enrollment, and experiments were conducted in accordance with the Declaration of Helsinki and approved by the institutional review boards at each site.

Clinical data and sample collection. Laboratory data, including UPCR, were collected within two weeks before kidney biopsy. Peripheral blood samples were acquired at the time of biopsy and at 3-, 6-, and 12-month postbiopsy. Samples were processed and stored at –80°C.

As previously described,[8] clinical responses were determined at 12 months using the Abatacept and Cyclophosphamide Combination: Efficacy and Safety Study trial definitions in patients with ISN/RPS class III, IV, V, or a combination thereof and a baseline UPCR ratio ≥1.0. Briefly, complete response (CR) was defined as a UPCR ≤ 0.5, normal serum creatinine or ≤125% of baseline if abnormal, and prednisone taper to 10 mg/day or less. Partial response (PR) was defined as the same creatinine criteria, a >50% reduction in UPCR, and a prednisone dose of ≤15 mg/day. Nonresponse did not meet the CR or PR definition.

Soluble mediators. Serum levels of TACE/ADAM17 (LifeTechnologies) were determined by enzyme-linked immunosorbent assay according to manufacturer instructions. Serum

Foundation, Bristol-Myers Squibb Company, Foundation for the National Institutes of Health, GlaxoSmithKline, Janssen Research and Development LLC, Lupus Foundation of America, Lupus Research Alliance, Merck Sharp & Dohme Corp, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer Inc, Rheumatology Research Foundation, Sanofi, and Takeda Pharmaceuticals International Inc) created to develop new ways of identifying and validating promising biologic targets for diagnostics and drug development. Supported by the NIH (grants UH2-AR-067676, UH2-AR-067677, UH2-AR-067679, UH2-AR-067681, UH2-AR-067685, UH2-AR-067688, UH2-AR-067689, UH2-AR-067690, UH2-AR-067691, UH2-AR-067694, UM2-AR-067678, UC2-AR-081039, UC2-AR-081032, P30-AR-073750, UM1-AI-144292, U01-AI-176244, and R01-DK-134625). Dr Fava's work was supported by the Jerome L. Greene Foundation, the Plank Foundation, and the Lupus Foundation of America (grant G-2104-01274).

[1]Johns Hopkins University, Baltimore, Maryland; [2]Oklahoma Medical Research Foundation, Oklahoma City; [3]New York University Grossman School of Medicine, New York; [4]Feinstein Institutes for Medical Research, Manhasset, New York; [5]Stanford University School of Medicine, Stanford, California; [6]Cedars-Sinai Medical Center, Los Angeles, California; [7]University of California San Francisco; [8]University of California San Diego; [9]Albert Einstein College of Medicine, Bronx, New York; [10]Bar-Ilan University, Zefat, Israel; [11]University of Rochester Medical Center, Rochester, New York; [12]Medical University of South Carolina, Charleston; [13]Northwell Health, Great Neck, New York; [14]Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [15]University of Oklahoma Health Sciences Center, Oklahoma City.

Drs Fava and Wagner are co-first authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25652).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25652.

Address correspondence via email to Judith A. James, MD, PhD, at jamesj@omrf.org.

Submitted for publication February 11, 2025; accepted in revised form August 21, 2025.

concentrations of an additional 66 analytes, including cytokines, chemokines, and tumor necrosis factor receptor (TNFR) superfamily members with published roles in LN and SLE (Supplementary Table 1), were analyzed by xMAP multiplex assays (Thermo Fisher Scientific), as previously described.[9] Data were acquired on the Bio-Rad BioPlex200 array system (Bio-Rad). The concentration (pg/mL) for each analyte was interpolated from five-parameter logistic regression modeling. Analytes below the detection limit were assigned a 0.001 pg/mL value. Analytes, in which over 95% of samples were below the range of detection (n = 7), were excluded from analyses (Supplemental Table 1).

Statistical analyses. A two-tailed Mann-Whitney U test was used to compare two groups, whereas the Kruskal-Wallis test with Dunn's multiple comparisons was applied for analyses involving multiple groups. Categorical variables were analyzed using the chi-square test. For multiple group comparisons across time points, statistical significance was determined using a two-way analysis of variance with Tukey's multiple comparisons. Correlations among soluble mediator concentrations and NIH-AI and NIH-CI were determined using Spearman's rank correlation. Linear trends over time were evaluated using a mixed-effects model. For multiple comparisons, *q* values were calculated using the two-stage step-up method of Benjamini, Krieger, and Yekutieli.

Baseline mediators associated with proliferative LN (class III or IV ± V) compared with nonproliferative LN (classes I, II, or V) were identified through bootstrapped least absolute shrinkage and selection operator (LASSO) regression (*glmnet* function in R). In total, 1,000 bootstrap iterations were performed to assess feature selection stability, retaining biomarkers selected in ≥60% of iterations. Firth's penalized logistic regression (*logistf* package in R) with 1,000 bootstrap iterations was used for univariate and multivariate analyses. Receiver-operating characteristic curves were analyzed using bootstrapped logistic regression (1,000 iterations; *pROC* package in R). Areas under the curve (AUC) were considered statistically significant if their 95% confidence intervals (CIs) excluded 0.5. Heatmaps and clustering using Ward's minimum variance method were generated using Complexheatmap version 4.2 in R. GraphPad Prism 9 was used for all other analyses. *P* and *q* values < 0.05 were considered statistically significant.

RESULTS

Serum soluble mediator expression by histologic class, activity, and chronicity. Serum soluble mediator profiling revealed widespread dysregulation in patients with LN (n = 268) compared with healthy controls (n = 22), with 34 mediators elevated and only tissue inhibitor of metalloproteinases 2 (TIMP-2) reduced (Supplemental Table 2; Supplemental Figure 1). Further analysis of soluble mediator expressions within LN histologic classes revealed distinct profiles (Supplemental Figure 2). Specifically, proliferative patients with LN (class III or IV ± V) exhibited high levels of several proinflammatory mediators, including syndecan-1, TNF receptor type I (TNFRI), TNF receptor type II (TNRFII), and vascular cell adhesion molecule 1 (VCAM-1), and lower expression of CCL3 or macrophage inflammatory protein 1α (MIP-1α), CD40 ligand (CD40L), and interleukin-5 (IL-5) compared with other LN classes (Figure 1A). Nonproliferative classes, including mesangial (class I or II), membranous (class V), and advanced sclerotic (class VI) LN, exhibited distinct mediator patterns (Figures 1B and 1C; Supplemental Figure 3). Because histologic class alone may not fully capture intrarenal inflammation or damage, we next assessed whether mediator levels correlate with activity and chronicity indices in a subset of patients with proliferative and membranous LN (n = 109). Soluble mediator concentrations were differentially associated with these indices (Figure 1D). Consistent with the higher activity characteristic of proliferative LN,[1] syndecan-1, TNFRI, and TNFRII were strongly associated with both histologic activity and chronicity, whereas VCAM-1 was associated with only activity (Figure 1D). In contrast, mediators inversely associated with proliferative LN, such as CD40L, were negatively correlated with histologic activity (Figure 1D). Notably, stem cell factor was uniquely associated with chronicity and advanced sclerosing LN and clustered with mediators positively associated with both activity and chronicity (Figure 1D; Supplemental Figure 3).

Serum soluble mediator profiles identify proliferative LN. Given that soluble mediator levels varied by histologic class and correlated with activity and chronicity indices, we next assessed whether soluble mediator levels could distinguish among proliferative patients (n = 160) and nonproliferative patients (classes I, II, or V; n = 96). Patients with advanced sclerosis (class VI) were omitted from the analysis because they represent an advanced, chronic stage of LN with little to no active inflammation.[1] Using bootstrapped LASSO regression on baseline levels of 60 soluble mediators (Supplemental Table 1), we identified 20 mediators stably associated with proliferative LN (selection frequency ≥60%; Supplemental Table 3). Of these, higher levels of syndecan-1, TNFRI, TNFRII, VCAM-1, CCL2/monocyte chemotactic protein 1, IL-6, and matrix metalloproteinase 3 (MMP-3), and lower levels of APRIL, CCL3/MIP-1α, CD40L, and IL-5 were most strongly associated with proliferative LN (Figure 2A and Supplemental Table 4). Notably, CCL2, CD40L, CXC ligand 8/IL-8, IL-6, IL-21, and MMP-3 were independently associated with proliferative LN in a multivariate model that included all the LASSO-selected features; only APRIL and CD40L remained significantly associated with proliferative LN following adjustment for extrarenal manifestations and clinical and/or demographic features associated with proliferative LN (Figure 2A; Supplemental Tables 4 and 5). Individually, some

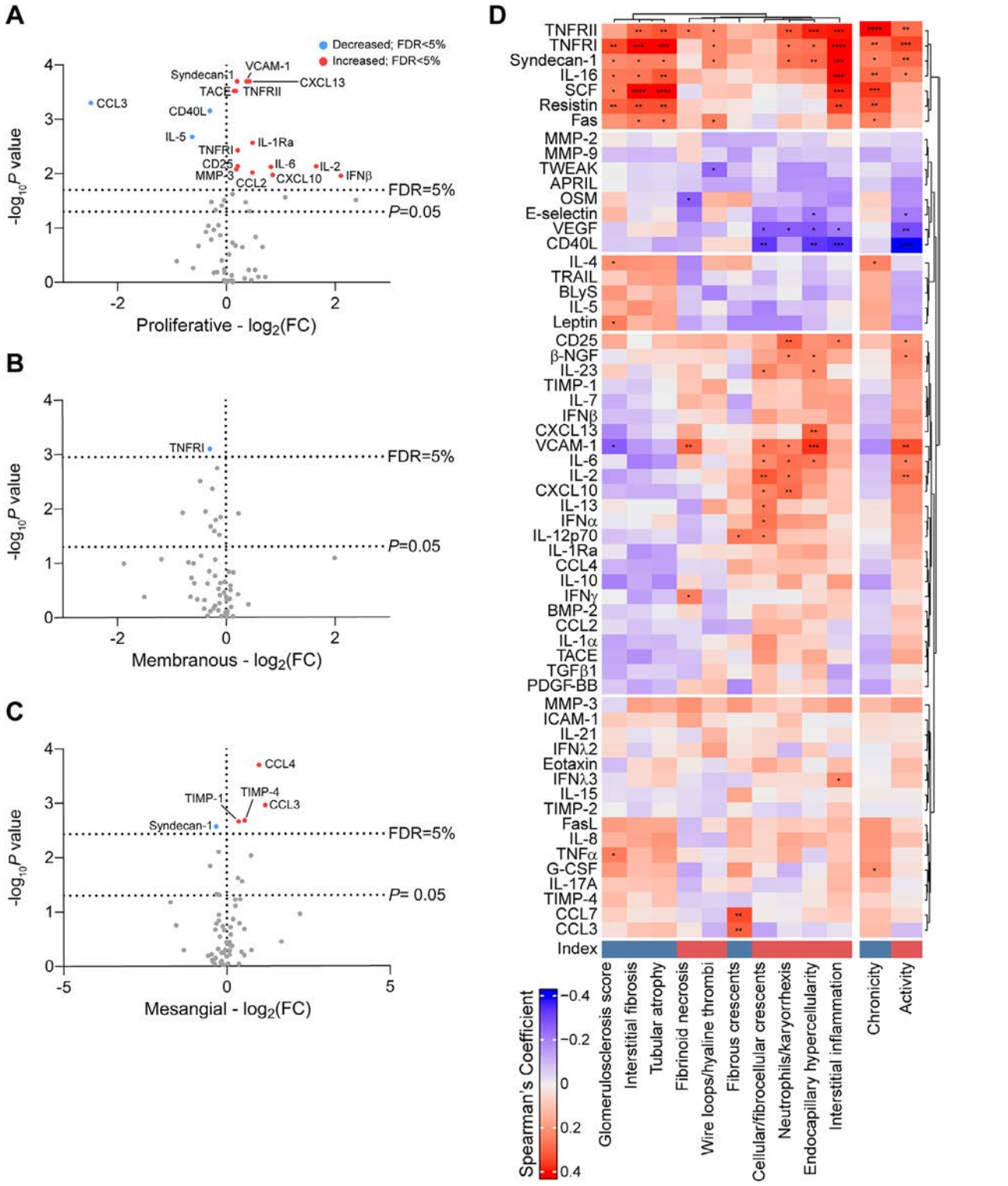

Figure 1. Serum soluble mediator expression differs by lupus nephritis histologic class, activity, and chronicity. (A–C) Volcano plots display the fold-change in mediator levels for (A) proliferative (class III or IV ± V; n = 160), (B) membranous (class V; n = 72), (C) and mesangial (class I or II; n = 24) classes compared with all other LN classes. Statistical significance was determined using a Mann-Whitney test and two-stage linear step-up procedure of Benjamini, Krieger, and Yekutieli. The horizontal dashed lines indicate the significance threshold. (D) Spearman's rank correlation coefficients for baseline soluble mediator concentrations and National Institutes of Health activity and chronicity indices and subindices (n = 109). Clustering was performed using Ward's minimum variance method. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. FC, fold change; FDR, false discovery rate; IFN, interferon; IL, interleukin; LN, lupus nephritis, MMP, matrix metalloproteinase; TNF, tumor necrosis factor; TNFRI, TNF receptor type I; VCAM-1, vascular cell adhesion molecule 1.

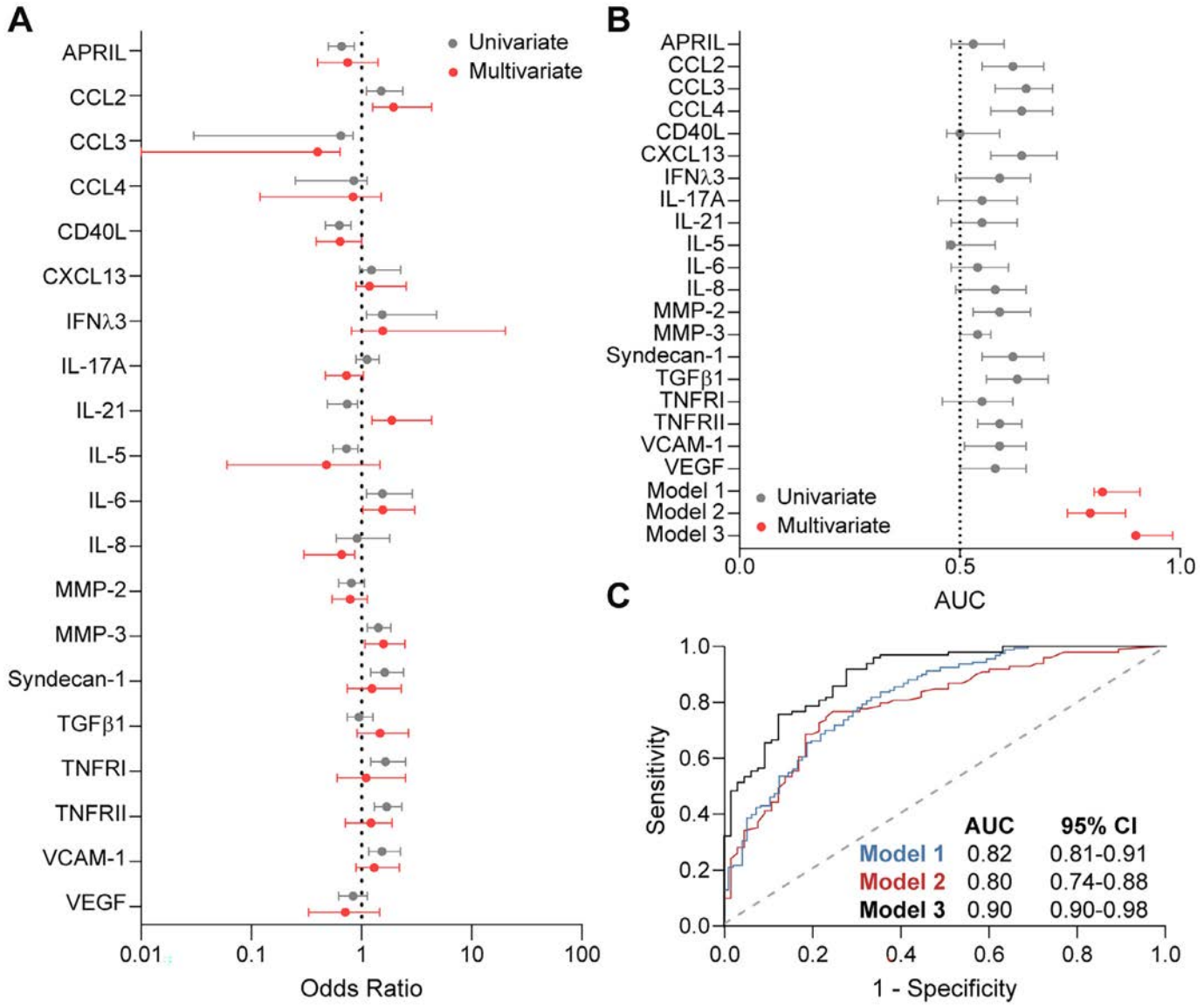

Figure 2. Serum soluble mediator profiles distinguish proliferative lupus nephritis. Soluble mediator concentrations were measured at the time of kidney biopsy in patients with proliferative (class III or IV ± V; n = 160) and nonproliferative (class I, II, or V; n = 96) lupus nephritis. Bootstrapped least absolute shrinkage and selection operator regression was used to select mediators associated with proliferative lupus nephritis. (A) Forest plot of associations among these least absolute shrinkage and selection operator–selected soluble mediator concentrations and proliferative lupus nephritis, as determined by univariate or multivariate logistic regression incorporating all selected mediators. Dots represent the odds ratio per 1 SD increase in concentration (Z score), and horizontal lines represent the 95% CI. (B) Forest plot displaying the AUC for the performance of least absolute shrinkage and selection operator–selected soluble mediator concentrations in predicting proliferative lupus nephritis. Dots represent the AUC, and horizontal lines represent the 95% CI. AUCs were considered statistically significant if the 95% CIs did not include 0.5 (dashed line). (C) Receiver-operating characteristic curves for multivariate logistic regression models predicting proliferative lupus nephritis: least absolute shrinkage and selection operator–selected soluble mediators (model 1), clinical or demographic factors (race, ethnicity, biopsy history, Systemic Lupus Erythematosus Disease Activity Index, C3, and C4; model 2), and combined least absolute shrinkage and selection operator–selected soluble mediators and clinical/demographic factors (model 3). AUC, area under the curve; CI, confidence interval; IFN, interferon; IL, interleukin; MMP, matrix metalloproteinase; TGF, transforming growth factor; VCAM-1, vascular cell adhesion molecule 1.

mediators showed modest predictive performance, with AUCs ranging from 0.59 to 0.65 (Figure 2B). In contrast, a model combining all LASSO-selected mediators achieved an AUC of 0.82 (95% CI 0.81–0.91; Figures 2B and 2C). The predictive performance was further improved when combined with demographic and clinical factors (AUC 0.90; 95% CI 0.90–0.98; Figures 2B and 2C). Importantly, the addition of soluble mediators outperformed demographic and clinical factors alone (AUC 0.80; 95% CI 0.74–0.88; Figure 2B and 2C), confirming the added value of these biomarkers over solely clinical data.

Early decreases in syndecan-1 and treatment response in proliferative patients with LN. Longitudinal analyses of proliferative patients with LN revealed that syndecan-1, TNFRI, TNFRII, and VCAM-1, mediators associated with proliferative LN and NIH activity, decreased significantly over 12 months in patients achieving CR (n = 25) or PR (n = 18), although the decrease in VCAM-1 was not statistically significant (Supplemental Figure 4; Supplemental Table 6). To evaluate early predictors of treatment response (CR or PR), we analyzed 3-month changes in proliferative LN-associated mediators. Reductions in syndecan-1 and VCAM-1 at 3 months, alongside declines in UPCR, were significantly associated with a 12-month treatment response in proliferative LN (Figure 3A; Supplemental Table 7). However, only changes in syndecan-1 remained independently predictive in multivariate analysis (odds ratio [OR] 0.004; 95% CI 0.000002–0.36; $P = 0.03$; Figure 3A; Supplemental Table 7). Induction therapy did not differ between responders and nonresponders (Supplemental Table 8). Syndecan-1 remained independently associated with treatment response following adjustment for clinical factors previously identified to associate with treatment response, including baseline UPCR < 3, >25% decrease in UPCR from baseline to week 12, chronicity index, and anti–double-stranded DNA positivity[10] (OR 0.002; 95% CI, 0.0000000001–0.07; $P = 0.003$; Supplemental Table 7). Receiver-operating characteristic analyses confirmed that reductions in syndecan-1 (AUC 0.89; 95% CI, 0.742–0.979) and VCAM-1 (AUC 0.81; 95% CI, 0.665–0.932) outperformed UPCR (AUC 0.78; 95% CI, 0.619–0.909) for predicting treatment response, and a combined model integrating syndecan-1, VCAM-1, and UPCR achieved maximal accuracy (AUC 0.9; 95% CI, 0.818–0.98; Figure 3B and 3C).

Molecular clustering of LN subgroups beyond histology. Having established that serum biomarkers can distinguish proliferative LN and predict treatment response, we next assessed whether broader molecular heterogeneity exists within LN. Unsupervised clustering of baseline soluble mediator profiles identified four distinct molecular clusters (Supplemental Figure 5A). Although Clusters 2 and 4 were enriched for proliferative LN (80% and 76%, respectively), patients of the same histologic class were distributed across multiple clusters, indicating that histology alone does not fully capture LN diversity (Supplemental Figure 5A; Supplemental Table 9). Molecular clusters also differed in clinical features, such as serum creatinine and complement levels, but treatment response rates and induction therapies were similar across clusters (Supplemental Table 9). Each cluster displayed a distinct soluble mediator signature (Supplemental Figures 5B–5E). For example, cluster 2, which had the highest chronicity numerically, was characterized by a broad down-regulation of multiple cytokines and growth factors compared with the other clusters (Supplemental Figure 5C). In contrast, cluster 3 featured elevated tissue remodeling proteins (TIMPs and MMP-9), proinflammatory cytokines (interferon [IFN] γ), and multiple chemokines (Supplemental Figure 5D), whereas cluster 4, which had numerically higher LN activity, was marked by increased type I IFNs and cytokines, including IL-12 and IL-23 (Supplemental Figure 5E). Together, these findings

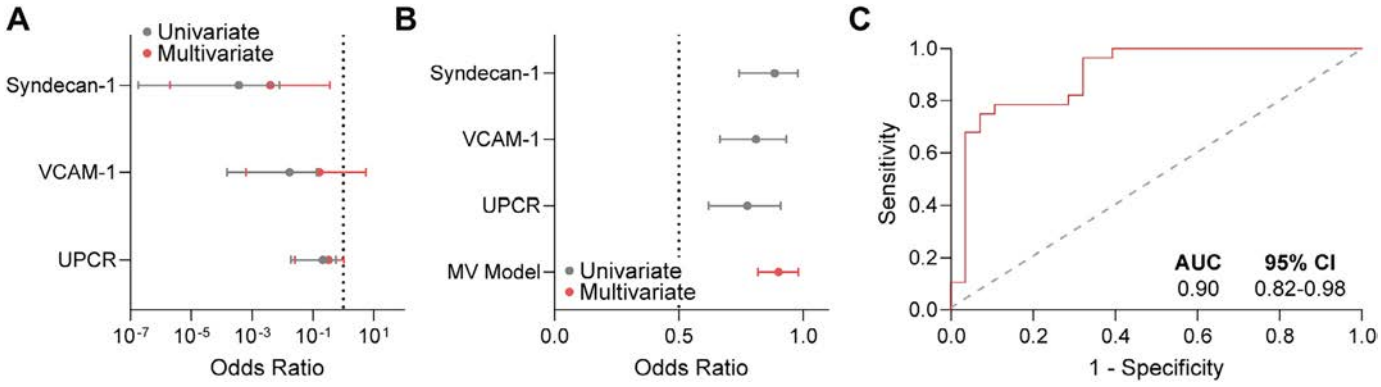

Figure 3. A subset of soluble mediators associated with proliferative lupus nephritis can predict treatment response in proliferative patients with lupus nephritis. Soluble mediator concentrations were measured at the time of kidney biopsy and 3 months after kidney biopsy in patients with proliferative lupus nephritis. Treatment response was determined at 12 months as complete response (n = 25), partial response (n = 18), or no response (n = 40). (A) Forest plot of significant associations between the decline at three months of LASSO-selected soluble mediator concentrations and any treatment response in proliferative patients with lupus nephritis, as determined by univariate or multivariate logistic regression, including all four mediators. Dots represent the odds ratio per 1 SD increase in concentration (Z score), and horizontal lines represent the 95% CI. (B) Forest plot displaying the AUC for the performance of changes in soluble mediator concentrations from baseline to three months in predicting any treatment response in patients with proliferative lupus nephritis. Dots represent the AUC, and horizontal lines represent the 95% CI. AUCs were considered statistically significant if the 95% CIs did not include 0.5 (dashed line). (C) Receiver-operating characteristic curve for the predictive performance of changes in syndecan-1, VCAM-1, and UPCR concentrations from baseline to 3 months. AUC, area under the curve; CI, confidence interval; MV, multivariate; UPCR, urine protein/creatinine ratio; VCAM-1, vascular cell adhesion molecule 1.

demonstrate substantial molecular heterogeneity within LN and among histologic classes, highlighting the potential of soluble mediator profiling as a "liquid biopsy" that could refine LN classification beyond standard histopathology by identifying key molecular pathways.

DISCUSSION

Several serum soluble mediators, including syndecan-1, TNFRII, and VCAM-1, have been previously associated with LN activity and chronicity.[5–7] However, few studies have investigated the potential of these biomarkers for distinguishing histologic class or early predictions of treatment response. Our study addresses this gap by using LASSO regression to identify 20 soluble mediators associated with proliferative LN, including syndecan-1, TNFRI, TNFRII, and VCAM-1, correlated with NIH-AI and NIH-CI. Notably, these four mediators declined over time in patients achieving CR or PR, suggesting a potential role in LN pathogenesis. Specifically, elevated levels of syndecan-1 and VCAM-1, both markers of cell adhesion and endothelial activation, along with TNFRI and TNFRII, may reflect ongoing vascular inflammation and tissue remodeling, critical drivers of LN progression. Increases in these mediators may also reflect underlying changes in proteolytic receptor shedding. TNFRI, TNFRII, and syndecan-1 are substrates of the disintegrin and metalloprotease TACE.[11] Although we did not observe differences in TACE concentrations in patients with LN compared with controls, its activity is regulated by complex mechanisms, such as interactions with the inactive rhomboid 2. As inactive rhomboid 2–dependent regulation of TACE is implicated in LN,[12] alterations in TACE activity may contribute to the inflammatory and tissue remodeling processes that drive LN, potentially influencing disease outcomes and treatment response.

Numerous large-scale proteomic studies have demonstrated the value of urinary biomarkers in LN for classifying and monitoring response to treatment. For example, reductions in urinary proteins that correlate with the NIH-AI, such as IL-16 and CD163, have been shown to decrease by 3 months in treatment responders and predict 1-year outcomes.[13] In addition, urinary IL-16 levels can differentiate patients with proliferative LN.[14] Notably, urinary VCAM-1 is also associated with multiple measures of LN disease activity, including UPCR.[15] In this study, reductions in serum concentrations of syndecan-1 or VCAM-1 at 3 months outperformed UPCR, the current standard of care, for predicting treatment response in proliferative LN. These findings support the expectation that decreases in inflammation result in a renal response. Notably, a combined biomarker panel of syndecan-1, VCAM-1, and UPCR increased the AUC to 0.9, demonstrating superior predictive power compared with individual markers alone. Integrating serum and urinary biomarkers could provide a more comprehensive approach to classifying LN disease activity and assessing treatment efficacy, capturing both inflammation and renal-specific injury. Future studies are needed to evaluate the potential of integrated biomarker panels in patients with LN to predict renal outcomes.

Beyond histologic class and conventional clinical features, our unsupervised clustering analysis revealed distinct molecular LN subgroups based on soluble mediator profiles. For example, a highly inflammatory cluster was characterized by elevated IL-12 and/or IL-23 and type I IFNs, suggestive of a myeloid-driven pathology, whereas another exhibited increased TIMPs, MMP-9, various chemokines, and proinflammatory cytokines, indicating active extracellular matrix remodeling and leukocyte recruitment. In contrast, we identified a cluster comprised solely of proliferative and membranous LN that was notable for higher chronicity and a broad downregulation of inflammatory mediators. This pattern may reflect a transition to a less inflamed, more fibrotic disease state, even within traditionally active histologic classes. This molecular heterogeneity, even among individuals with similar histology or clinical response, suggests the potential for integrating molecular subtyping with clinical data to refine risk stratification, enable biomarker-guided reclassification, and inform pathway-targeted therapeutic trials in LN. Future studies should validate the clinical use of these clusters, linking them to long-term outcomes and testing their value for precision treatment selection.

This study has some limitations. Because our study used a real-world cohort, it lacked the standardized treatment protocols of a clinical trial; thus, there was variability in dosing and induction therapy, reflecting routine clinical care. However, induction therapies did not differ based on treatment response in our cohort. Our cohort did not include other inflammatory or kidney diseases; thus, whether these findings are specific to LN is unclear. Some markers were outside the limits of detection and could not be accurately quantified. Finally, our study did not include multiyear follow-up to assess long-term outcomes, such as progression to end-stage renal disease.

In conclusion, this study identifies a panel of serum soluble mediators that may serve as noninvasive, longitudinal biomarkers for monitoring proliferative LN. Specifically, reductions in syndecan-1 and VCAM-1 at 3 months are associated with and predictive of 1-year treatment response to standard of care in proliferative patients with LN. Monitoring these biomarkers over time could enhance clinical management by informing treatment decisions and identifying potential therapeutic targets. In addition, molecular heterogeneity exists within LN, supporting the potential for biomarker-driven reclassification and pathway-specific therapeutic approaches.

ACKNOWLEDGMENTS

This text has been reviewed for spelling and grammar using a large language model.

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 James 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. Weening JJ, D'Agati VD, Schwartz MM, et al. The classification of glomerulonephritis in systemic lupus erythematosus revisited. Kidney Int 2004;65(2):521–530.
2. 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.
3. Lu J, Tam L-S, Lai FM-M, et al. Repeat renal biopsy in lupus nephritis: a change in histological pattern is common. Am J Nephrol 2011;34(3): 220–225.
4. Mok CC, Teng YKO, Saxena R, et al. Treatment of lupus nephritis: consensus, evidence and perspectives. Nat Rev Rheumatol 2023; 19(4):227–238.
5. Wu T, Ding H, Han J, et al. Antibody-array-based proteomic screening of serum markers in systemic lupus erythematosus: a discovery study. J Proteome Res 2016;15(7):2102–2114.
6. Yu KY, Yung S, Chau MK, et al. Clinico-pathological associations of serum VCAM-1 and ICAM-1 levels in patients with lupus nephritis. Lupus 2021;30(7):1039–1050.
7. Yu KYC, Yung S, Chau MKM, et al. Serum syndecan-1, hyaluronan and thrombomodulin levels in patients with lupus nephritis. Rheumatology (Oxford) 2021;60(2):737–750.
8. Fava A, Wagner CA, Guthridge CJ, et al. Association of autoantibody concentrations and trajectories with lupus nephritis histologic features and treatment response. Arthritis Rheumatol 2024;76(11):1611–1622.
9. Munroe ME, Young KA, Kamen DL, et al. Discerning risk of disease transition in relatives of systemic lupus erythematosus patients utilizing soluble mediators and clinical features. Arthritis Rheumatol 2017; 69(3):630–642.
10. Izmirly PM, Kim MY, Carlucci PM, et al. Longitudinal patterns and predictors of response to standard-of-care therapy in lupus nephritis: data from the Accelerating Medicines Partnership lupus network. Arthritis Res Ther 2024;26(1):54.
11. Kawai T, Elliott KJ, Scalia R, et al. Contribution of ADAM17 and related ADAMs in cardiovascular diseases. Cell Mol Life Sci 2021;78(9): 4161–4187.
12. Qing X, Chinenov Y, Redecha P, et al. iRhom2 promotes lupus nephritis through TNF- α and EGFR signaling. J Clin Invest 2018;128(4): 1397–1412.
13. Fava A, Buyon J, Magder L, et al. Urine proteomic signatures of histological class, activity, chronicity, and treatment response in lupus nephritis. JCI Insight 2024;9(2):e172569.
14. Häyry A, Faustini F, Zickert A, et al. Interleukin (IL) 16: a candidate urinary biomarker for proliferative lupus nephritis. Lupus Sci Med 2022; 9(1):e000744.
15. Kiani AN, Wu T, Fang H, et al. Urinary vascular cell adhesion molecule, but not neutrophil gelatinase-associated lipocalin, is associated with lupus nephritis. J Rheumatol 2012;39(6):1231–1237.

APPENDIX A: ACCELERATING MEDICINES PARTNERSHIP RHEUMATOID ARTHRITIS AND SYSTEMIC LUPUS ERYTHEMATOSUS NETWORK

Accelerating Medicines Partnership in RA/SLE Network contributors are as follows: Chun-Hao Lee, Derek Fine, Manny Monroy-Trujillo, Ummara Shah, Mariko Ishimori, Robert M. Clancy, Michael Belmont, Ming Wu, Nicole Bornkamp, Evan Der, Beatrice Goilav, Nicole Jordan, Daniel Schwartz, James Pullman, David Wofsy, Dawn Smilek, Patti Tosta, Matthias Kretzler, Celine C. Berthier, E. Steve Woodle, Dave Hildeman, and Michael Brenner.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 528–536
DOI 10.1002/acr.25657

American College of Rheumatology
Empowering Rheumatology Professionals

Ultrasonography Compared With Computed Tomography in Identifying Thoracic Aortic Aneurysms in Patients With Giant Cell Arteritis

Anne C. Bull Haaversen,[1,2] Lene K. Brekke,[3] Tanaz A. Kermani,[4] Øyvind Molberg,[2,5] and Andreas P. Diamantopoulos[6]

Objective. The objectives of this study were to evaluate the correlation and agreement between ultrasonography and computed tomography (CT) in measuring ascending aorta diameter in patients with giant cell arteritis (GCA) and to investigate the development of new ascending aortic aneurysms in patients with newly diagnosed GCA.

Methods. To achieve the primary objective, we performed a cross-sectional study that included patients with GCA who were evaluated with both ultrasonography and CT of the ascending aorta within six months of each other. For the second objective, patients with baseline aortic imaging at GCA diagnosis were observed prospectively with new aortic imaging at the end of the follow-up. A cutoff of ≥40 mm defined an aneurysm in both the ascending and the descending aorta.

Results. The cross-sectional study included 140 patients with a total of 169 comparable CT and ultrasonography scans and a mean disease duration of 47.8 months (95% confidence interval [CI] 42.1–53.5). A strong positive correlation, as measured by Spearman's rho ($r = 0.801$, $P < 0.001$), and a strong agreement, as indicated by an intraclass correlation coefficient of 0.849 ($P < 0.001$), were found between ultrasonography and CT. In the prospective follow-up study, 40 patients were included. Seven (17.5%) patients developed aneurysms in the ascending aorta in a mean follow-up time of 54 months (95% CI 43.7–64.6); all of the patients experienced large vessel (LV) involvement (two patients with LV-GCA and five patients with mixed-GCA).

Conclusion. A strong correlation and agreement were observed between ultrasonography and CT measurements in assessing the diameter of the ascending aorta in patients with GCA. Ultrasonography may be a valuable tool for longitudinal monitoring and screening for aneurysm development. About 18% of patients developed aortic aneurysms over a mean follow-up of 4.5 years, underscoring the critical need for regular aortic screening in this population.

INTRODUCTION

Giant cell arteritis (GCA) is a granulomatous vasculitis that involves the aorta and its branches.[1] Patients with GCA are known to have a 2– to 17–fold increased risk of developing a thoracic aortic aneurysm compared with the general population,[2–6] with no or only a slightly increased risk for abdominal aortic aneurysms.[2,3,6] A meta-analysis of studies reported a pooled prevalence of 15% of thoracic aortic aneurysms.[7] However, the heterogeneity among individual studies was substantial, and diverging estimates may reflect the different populations studied, whether patients were screened systematically, and various methods used for both defining and diagnosing an aneurysm.

Aortic dilatations appear to be a late complication of GCA; the incidence of aortic aneurysms increases five years after the diagnosis of GCA and continues developing over time.[8,9] A recently published prospective study suggests that the intensity and extent of initial inflammation, as evaluated by positron emission tomography/computed tomography (CT) at diagnosis, may be associated with the risk of subsequent aortic dilatation.[10]

Previous studies indicate that the risk mainly involves the ascending part of the thoracic aorta.[2,5,11–14] Thus, ultrasonography

[1]Martina Hansens Hospital, Bærum, Norway; [2]University of Oslo, Oslo, Norway; [3]Hospital for Rheumatic Diseases, Haugesund, Norway; [4]University of California, Los Angeles; [5]Oslo University Hospital, Rikshospitalet, Oslo, Norway; [6]Akershus University Hospital, Lørenskog, Norway.

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.25657).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25657.

Address correspondence via email to Anne C. Bull Haaversen, MD, at annecbull@gmail.com.

Submitted for publication April 9, 2025; accepted in revised form September 4, 2025.

SIGNIFICANCE & INNOVATIONS

- The first study to evaluate ultrasonography and computed tomography for ascending aortic measurements in giant cell arteritis (GCA).
- This study demonstrates strong correlation and agreement between the two modalities, supporting ultrasonography as a follow-up tool regarding aortic monitoring in GCA.
- Prospective data show that 18% of patients with GCA develop ascending aortic aneurysms within 4.5 years of diagnosis.

evaluation of the ascending aorta could be helpful in screening for aortic aneurysms in patients with GCA.

It is unclear how many patients need to be screened to capture one aneurysm requiring surgery or to prevent other aneurysm-specific interventions. In a previous systematic review, five to ten patients with GCA would require aortic imaging to detect one previously unknown thoracic aortic aneurysm or dilation.[7]

Guidelines recommend imaging the aorta and its branches at the time of GCA diagnosis to identify large vessel (LV) involvement.[15–19] However, consensus regarding the optimal imaging modalities or frequency of screening the aorta for aneurysms is still lacking. Given that aneurysms can develop several years after the diagnosis and also in patients considered to be in long-term remission, it is also unclear how long the monitoring should continue.[2,8,9] EULAR recommends screening patients with signs and symptoms of vascular damage, as well as those with persistent inflammation of large arteries, including the aorta. The research agenda proposes to clarify these aspects to prevent complications such as aortic dissection ruptures.[17] Furthermore, the current American College of Rheumatology (ACR) guidelines recommend long-term clinical monitoring, which may include imaging studies; however, they do not specify the optimal imaging modality or screening interval.[19]

De Boysson et al suggest annual echocardiography to monitor the aorta.[5] Compared with CT, echocardiography may be a more feasible imaging method because it is cost-effective and patient-friendly, with no radiation and no use of contrast. In Norway, echocardiography is already the primary method for observing patients with ascending aortic aneurysms in general. CT is reserved for monitoring aneurysms elsewhere in the aorta, such as the descending aorta, which ultrasonography cannot adequately evaluate.

No studies have compared the different imaging modalities for detecting aneurysms in patients with GCA. However, several studies in cardiology comparing ultrasonography with CT have demonstrated good agreement and correlation.[20,21]

The primary objective of the cross-sectional study was to evaluate the correlation and agreement between ultrasonography and CT measurements of the ascending aorta in patients with GCA. The second objective was to investigate new aneurysm formation in the ascending aorta in a prospective cohort of patients with GCA.

MATERIALS AND METHODS

Patient selection. The study included patients with GCA observed at the Department of Rheumatology, Martina Hansens Hospital in Bærum, Norway. The diagnosis was based on typical clinical manifestations and imaging (CT or ultrasonography), and all the patients had a definite diagnosis of GCA according to the proposed extension of the ACR 1990 classification criteria.[22,23] All patients in the cross-sectional part diagnosed after 2017 had positive ultrasonography findings of the temporal arteries and/or LVs, whereas patients diagnosed before 2017 had either a positive temporal artery biopsy or positive LV imaging by CT scan. Patients with GCA who underwent CT and/or ultrasonography imaging of the ascending aorta at baseline and follow-up were included in the prospective part of this study to enable direct comparison and longitudinal assessment of aortic diameter changes. The collected data included demographics, imaging, and clinical findings. Patients were categorized into the following three GCA subtypes: cranial-GCA (c-GCA), LV-GCA, and mixed-GCA. The aorta was assessed between September 2017 and June 2024.

Imaging methods and definitions. The ascending aorta was examined using a Canon Aplio 800 ultrasonography machine with a 2.5 MHz phased-array transducer. All ultrasonography evaluations of the LVs were performed using the settings recommended by Terslev et al.[24] The aortic dimension was evaluated at the parasternal long-axis view as the maximal diameter of the proximal ascending aorta, approximately 1 cm after the sinotubular junction, using the inner edge to inner edge technique, which aligns with standard practice in CT-based aortic diameter assessment. Measurements were not timed to the cardiac cycle. The cut-off value of ≥40 mm, defining an aneurysm in the ascending aorta, was applied to all patients regardless of age, sex, height, or body surface area.[25–28] The ultrasonography examinations were performed by experienced sonographers (APD, ACBH). APD, who performed most of the ultrasonography examinations, was fully blinded to the CT results for all patients. ACBH was blinded for the majority of cases, with only a minority of examinations performed unblinded because of logistical constraints. Extensive training was provided in collaboration with a cardiologist experienced in echocardiography, combining theoretical guidance and hands-on sessions with models and patients to develop and refine the ultrasonography protocol.

There was no central measurement of aortic diameter. Measurements were obtained locally by experienced radiologists and sonographers following standardized protocols at each center.

The CT scans were performed in three different centers without a standardized acquisition protocol. No calibration was performed between centers or imaging protocols. The referral requested measurement of the ascending aorta, consistent with routine clinical practice for monitoring these patients. The decision to use contrast during CT scans was made by the radiologists based on individual patient conditions and comorbidities, such as kidney function. Consequently, the use of intravenous contrast varied across the cohort. We accepted both contrast-enhanced CT (CT angiography [CTA]) and non-contrast CT scans for aortic assessment. This is justified because contrast is not required for an accurate aortic diameter; non-contrast CT demonstrates measurement agreement with CTA at the millimeter level.[29] Additionally, dedicated subgroup analyses were performed for aortic CTA and non-contrast CT.

Image interpretation was performed by different radiologists, reflecting real-world clinical practice. The maximum diameter of the ascending aorta was reported. A cut-off value of ≥40 mm defined an aneurysm in both the ascending and descending aorta.

Cross-sectional study. Ultrasonography and CT imaging were performed within six months of each other at any point in the GCA disease course.

Prospective follow-up study. The baseline imaging, either CT or ultrasonography, was performed within six months of diagnosis. Follow-up imaging, either CT or ultrasonography, was performed at various times. When both CT and ultrasonography were performed at baseline and/or follow-up, the highest measurement value of the ascending aorta was used in the analyses.

Statistical analysis. Continuous parameters were expressed as means with SDs, and non-normally distributed parameters were expressed as medians with interquartile ranges (IQRs). Categorical parameters were presented as frequencies and percentages. Categorical variables were compared using Chi-square or Fisher's exact tests. Normality of continuous variables was assessed using the Shapiro-Wilk test and Q-Q plots. Variables with normal distribution were analyzed using one-way analysis of variance, whereas non-normally distributed variables were compared using nonparametric tests (Mann–Whitney or Kruskal–Wallis).

To assess the agreement between ultrasonography and CT measurements of the ascending aorta, we calculated the intraclass correlation coefficient (ICC) using a two-way mixed-effects model with absolute agreement definition and the Bland–Altman analysis. ICC values were reported with 95% confidence intervals (CI). For the Bland–Altman analysis, the difference between paired measurements (CT minus ultrasonography) was plotted against their mean for each patient. The mean difference was calculated to evaluate systematic bias, and the 95% limits of agreement were defined as the mean difference ± 1.96 times the SD of the difference. In addition, Spearman's rank correlation coefficient was used to assess the association between the two modalities.

Comparisons between patients with and without ascending aorta dilatation (≥40 mm) were first performed using univariable analyses (for continuous variables, the Mann–Whitney U test was used, and for categorical variables, Fisher's exact test was used). We conducted a multivariable binary logistic regression analysis to identify factors associated with ascending aorta dilatation. The dependent variable was the presence of ascending aorta ≥40 mm. Independent variables included age at diagnosis, sex, LV involvement, and GCA subtype. Odds ratios (ORs) with 95% CI were calculated. Model fit was assessed using the Hosmer–Lemeshow goodness-of-fit test. Time-to-event analysis, with aortic aneurysms as the event, was performed using Kaplan-Meier survival analysis. Patients with documented aneurysms already at the time of GCA diagnosis were excluded from the time-to-analysis.

Statistical analyses were performed using SPSS, version 21 (SPSS Inc, Chicago, IL). *P* values less than 0.05 were considered significant.

Ethical considerations. According to the Declaration of Helsinki, written informed consent was obtained from all participating patients. The Regional Ethics Committee approved the study (reference number: REK Sør-Øst 2019/1362). Given the limited sample size and the scope of analyses performed, this study should be considered exploratory.

RESULTS

Cross-sectional study. A total of 140 patients with GCA, 88 (62.9%) who were female with a mean age of 71.1 years (SD 7.8), were included in the cross-sectional study. Forty-eight (34.3%) patients had c-GCA, 27 (19.3%) had LV-GCA, and 65 (46.4%) had mixed-GCA. The mean follow-up time from diagnosis to study scans (the last aorta imaging) was 47.8 months (95% CI 42.1–53.5). All 140 patients had at least 1 comparable CT and ultrasonography scan, 27 patients had 2 comparable scans, and 2 patients had 3 comparable scans, resulting in a total of 169 scans. Non-contrast CT was used in 50 examinations, whereas CTA was used in 119 examinations. The median diameter was 37 mm for non-contrast CT (n = 50) and 35 mm for CTA (n = 119), yielding a median difference of 2 mm (P = 0.004). Although statistically significant, this difference is well below the accepted threshold of 5 mm, which has been consistently identified as the limit beyond which changes exceed measurement variability and become clinically meaningful.[30]

Correlation and agreement assessment of ultrasonography and CT measurements of the ascending aorta. The median diameter of the ascending aorta was 34 mm (IQR 5) as measured by ultrasonography and 36 mm (IQR 7) as measured by CT (Figure 1). Spearman's rho demonstrated a strong positive correlation between ultrasonography and CT measurements (r = 0.801, 95% CI 0.78–0.89, P < 0.001).

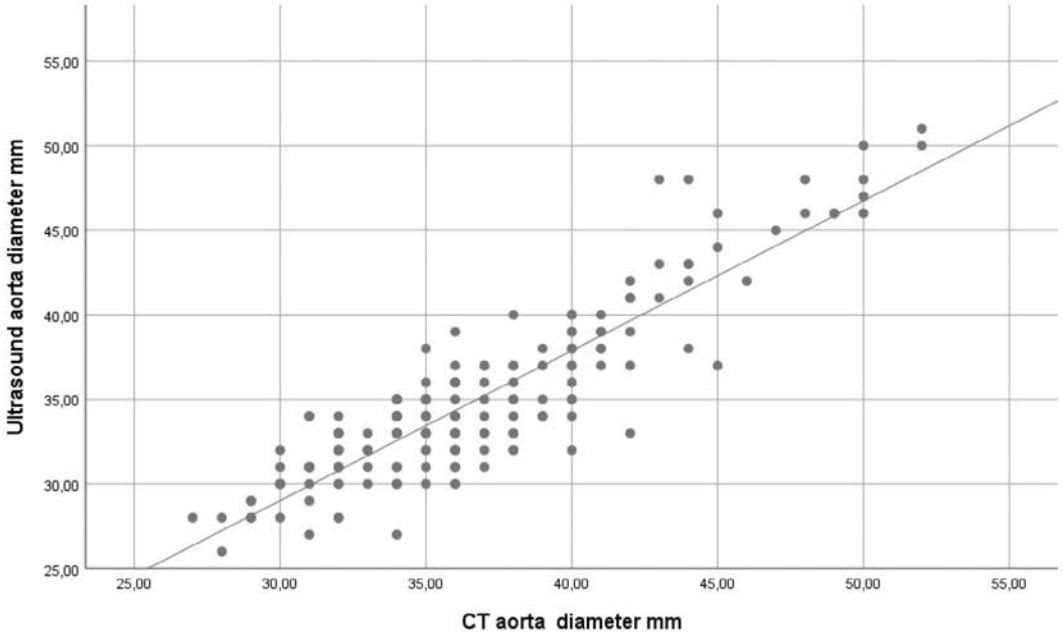

Figure 1. Measurements of the diameter of the ascending aorta by ultrasonography and CT (mm). CT, computed tomography.

Agreement was further assessed using the ICC, which was based on a two-way mixed-effects model with absolute agreement. The single measures ICC was 0.850 (95% CI 0.578–0.928, $P < 0.001$), indicating excellent agreement between modalities. The Bland–Altman analysis revealed a mean difference of 1.8 mm (SD 2.37 mm) between CT and ultrasonography measurements, indicating that ultrasonography slightly underestimates the aortic diameter compared with CT. The 95% limits of agreement ranged from –2.88 mm to 6.40 mm (Supplementary Figure 1). The agreement remained excellent, as evidenced by the inclusion of only patients with CTA in the analysis (ICC = 0.877, 95% CI 0.827–0.913).

Prevalence of aneurysm in the ascending aorta. Ascending aortic aneurysms were visualized in 37 of 140 patients (26.4%), with a median disease duration of 25 months (IQR 59). The baseline data of the patients with GCA in the cross-sectional study are presented in Table 1. Univariable analysis demonstrated that male sex was significantly more common in patients with ascending aortic aneurysm ($P < 0.001$), whereas no association was seen for age, GCA subtype, or LV involvement. In binary logistic regression analysis, sex was the only independent predictor of ascending aortic aneurysm, with males having over four times increased odds (OR 4.19, 95% CI 1.90–9.25, $P < 0.001$). Age, GCA subtype, and LV involvement were not found to be associated with aneurysm development. The model demonstrated acceptable overall fit ($P = 0{,}251$). Of the 92 patients with LV involvement, 28 (30.4%) had an aneurysm, and among the 48 patients with c-GCA, 9 (18.8%) had an aneurysm ($P = 0.14$). None of the aneurysms required surgery, and there were no dissections seen on CT.

Six CT scans (4%) revealed a descending aortic aneurysm; however, five CT scans also had a coexisting ascending aortic aneurysm, and only 1 (0.7%) had an isolated descending aortic aneurysm (diameter of the descending aorta: 41 mm; diameter of the ascending aorta: 36 mm).

Prospective follow-up study. In the prospective study, 50 patients with GCA were eligible for inclusion based on the availability of aortic imaging, either by CT or ultrasonography, at both baseline and the end of follow-up.

Among the 50 patients, 10 (20%) had an ascending aortic aneurysm already at baseline and were therefore excluded from the analyses (two patients with c-GCA, two patients with LV-GCA, and six patients with mixed-GCA). Among the excluded 10 patients, the median diameter of the ascending aorta, which was measured by CT (except for two patients, for whom ultrasonography was used only), was 41 mm (IQR 2.75). Thus, 40 patients were included in the prospective follow-up study. At

Table 1. Baseline characteristics of the cross-sectional study cohort*

Baseline characteristics	All patients (n = 140)
Female, n (%)	88 (62.9)
Age, mean (95% CI), y	71.5 (69.5–73.4)
Subtype GCA, n (%)	
c-GCA	48 (34.3)
LV-GCA	27 (19.3)
Mixed-GCA	65 (46.4)
LV involvement	92 (65.7)

* CI, confidence interval; c, cranial; GCA, giant cell arteritis; LV, large vessel.

Table 2. Baseline characteristics of the patients included in the follow-up study (n = 40)*

Baseline characteristics	c-GCA (n = 6)	Mixed-GCA (n = 26)	LV-GCA (n = 8)	*P* value
Age at diagnosis, mean (95% CI), y[a]	70 (63.4–76.6)	69.1 (65.1–73.2)	67.4 (60.9–73.9)	0.52
Female, n (%)[b]	3 (50.0)	14 (53.8)	7 (87.5)	0.20
Duration follow-up months, median (IQR)[c]	40 (19)	49 (21)	52 (6)	0.31
Symptoms[b]				
Constitutional, n (%)	5 (83.3)	19 (73.1)	4 (50.0)	0.51
Headache, n (%)	5 (83.3)	17 (65.4)	4 (50)	0.60
Vision loss, n (%)	0	1 (3.8)	0	0.77
Polymyalgic symptoms, n (%)[b]	2 (33.3)	8 (30.8)	2 (25)	0.98
Jaw claudication, n (%)[b]	3 (50)	13 (50)	1 (12.5)	0.20
CRP, median (IQR)[c]	156 (107)	113 (92)	100 (86)	0.51
Ultrasonography evaluating ascending aorta, n (%)[b]	3 (50)	16 (61.5)	2 (25)	0.19
CT evaluating ascending aorta, n (%)[b]	3 (50)	21 (80.8)	8 (100)	0.07
Ultrasonography and CT evaluating ascending aorta, n (%)[b]	0 (0)	11 (42.3)	2 (25)	0.12

* ANOVA, analysis of variance; c-GCA, cranial giant cell arteritis; CRP, C-reactive protein; CT, computed tomography; IQR, interquartile range; LV-GCA, large vessel–giant cell arteritis.
[a] ANOVA.
[b] Fisher's exact test.
[c] Kruskal-Wallis.

baseline, 32 (80%) patients underwent a CT scan, and 21 (52.5%) underwent an ultrasonography. All 40 patients underwent an ultrasonography at follow-up, and 35 (87.5%) also received a CT scan. The highest aorta measurements were chosen for the analyses when both ultrasonography and CT were performed. The baseline characteristics of the patients included are seen in Table 2.

Development of aneurysm in the thoracic aorta during follow-up. Among the 40 patients included in the analysis, 7 (17.5%) developed aneurysms in the ascending aorta during a mean follow-up period of 54 months (95% CI 43.7–64.6) (*P* = 0.75), as shown in Figure 2 (95% CI 71.8–74.5). Of these 7 patients, 2 (28.6%) had LV-GCA, and 5 (71.4%) had mixed-GCA (*P* = 0.44). The characteristics of patients with aneurysms are presented in Table 3. None of the 35 patients (87.5%) with a follow-up CT of the aorta developed an aneurysm in the descending aorta.

Changes in the diameter of the ascending aorta during follow-up. The mean diameter of the ascending aorta at baseline was 33.0 mm (95% CI 32.0–34.0), whereas the mean diameter

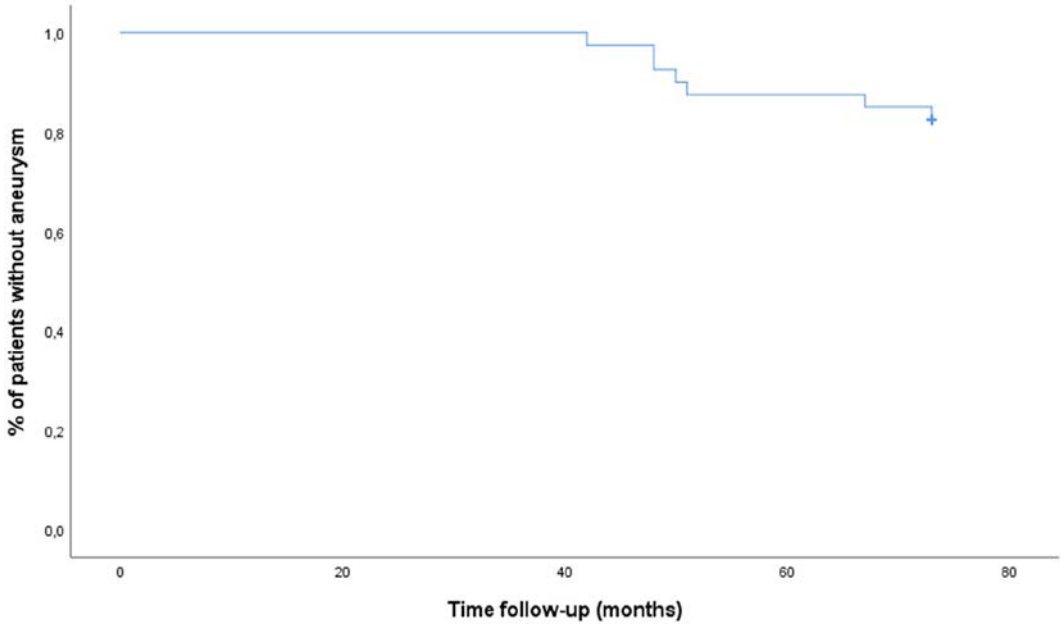

Figure 2. Kaplan-Meier survival plot for aneurysm development in the ascending aorta.

Table 3. Characteristics of the patients with aneurysm in the ascending aorta at follow-up*

	GCA subtype at diagnosis	Sex	Age at diagnosis, years	Ascending aorta diameter at diagnosis, mm, CT	Ascending aorta diameter at diagnosis, mm, ultrasonography	Ascending aorta diameter at follow-up, mm, CT	Ascending aorta diameter at follow-up, mm, ultrasonography	The time between GCA diagnosis and last aorta imaging, months
Patient 1	Mixed-GCA	Male	61	N/A	34	40	34	51
Patient 2	Mixed-GCA	Male	52	36	N/A	38	40	67
Patient 3	Mixed-GCA	Female	71	35	N/A	40	39	73
Patient 4	LV-GCA	Female	62	36	N/A	50	50	50
Patient 5	Mixed-GCA	Female	61	37	N/A	N/A	40	48
Patient 6	LV-GCA	Male	72	38	39	40	37	48
Patient 7	Mixed-GCA	Male	64	37	33	40	38	42

* CT, computed tomography; LV-GCA, large vessel–giant cell arteritis; N/A, not available.

at follow-up was 35.6 mm (95% CI 34.3–36.8) ($P < 0.01$) (Figure 3) (Mean difference: 2.6 mm, 95% CI 1.7–3.4, $P < 0.01$).

DISCUSSION

There is an unmet need for an applicable imaging method in the follow-up of patients with LV involvement in GCA. CT is considered the gold standard but has several drawbacks, including radiation exposure and limited availability. This study demonstrated a strong positive correlation between ultrasonography and CT measurements of the ascending aorta in patients with GCA, as well as a strong positive agreement (ICC and Bland–Altman tests). These results suggest acceptable correlation and agreement between ultrasonography and CT, with a modest but consistent bias toward lower ultrasonography measurements. Yet, the underestimation of 1.8 mm is low in terms of clinical significance for such minor differences.

According to European Society of Cardiology guidelines, transthoracic echocardiography (TTE) is recommended as the first-line imaging technique in evaluating thoracic aortic diseases. It can be used for the follow-up of aneurysms; however, if there is an increase of ≥3 mm per year in aortic diameter by TTE, confirmation by cardiovascular CT or magnetic resonance imaging (MRI) should be considered.[28]

To the best of our knowledge, no previous study has assessed the correlation and agreement of CT and ultrasonography measurements of the aorta during the follow-up of patients with GCA, even though some have suggested

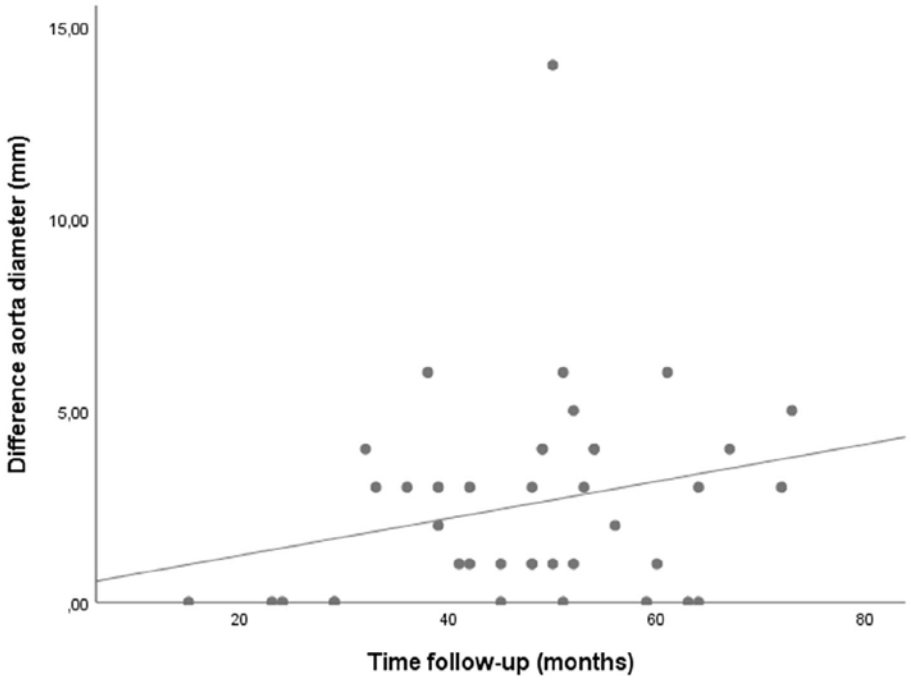

Figure 3. Change in diameter of the ascending aorta from baseline to follow-up. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25657/abstract.

echocardiography as a potential scanning tool.[4,5] Studies from cardiologists have demonstrated excellent inter- and intraobserver variability using TTE, CT, and MRI; however, there is a significant underestimation of all aortic diameters assessed by echocardiography.[21]

In the prospective cohort, 20% already had an ascending aortic aneurysm at diagnosis. In the 40 patients with GCA without aortic abnormalities at diagnosis, imaging with CT or ultrasonography demonstrated that 17.5% developed an aneurysm in the ascending aorta during a mean follow-up time of 4.5 years. In earlier studies of aneurysms in patients with GCA, the prevalence and incidence had been diverging, presumably because of the different populations studied and various methods used for defining and diagnosing an aneurysm. The studies had a variety of aneurysm definitions, imaging methods, and time points.[2,3,5–7,11] A meta-analysis of studies reported a pooled prevalence of 15% of thoracic aortic aneurysms.[7] Subsequent retrospective studies estimated that 10% to 14% of patients with GCA develop aortic aneurysms, mainly in the thoracic segment.[5,9,31,32] In the cross-sectional study, the prevalence was higher, at 26.4%, after a median disease duration of 25 months. An explanation for this discrepancy may be that in the retrospective studies, the definition of an aortic aneurysm was usually based on diagnostic codes from registers or radiologic reports without clear cutoffs being reported or used. Additionally, patients were not routinely screened, which may have underestimated the risk of developing aneurysms.[2,31,32] In contrast, a higher rate of patients (22%–33%) developed aortic aneurysms in prospective studies. Our results align with another cross-sectional study with 54 patients that demonstrated aortic aneurysm in 22.2%, with a median disease duration of 5.4 years.[33] In another study, 36 patients were prospectively observed for an extended period, with a median of 10.3 years. Twelve of these patients (33.3%) developed an aneurysm in the thoracic aorta.[8]

Of the 140 patients in the cross-sectional study, 6 patients had an aneurysm in the descending aorta, as visualized by CT. Yet, only one of these patients did not have a simultaneous aneurysm in the ascending aorta (dilatation of 36 mm in the ascending aorta, dilatation of 41 mm in the descending aorta and aortic arch, as shown by CT). Our findings confirm that thoracic aneurysms mainly occur in the ascending part of the aorta, which is consistent with recent published studies.[8,33] Older imaging studies of the aorta did not specify the location of the aneurysm within the thoracic aorta. An explanation for most aneurysms in the ascending part of the aorta could be that aortic dilatation has been demonstrated using various imaging tools, including echocardiography.[34] One study demonstrated that most thoracic aneurysms are located in the descending part; however, in this study, the prevalence of coexisting aneurysms in the ascending part was not reported.[10] In addition, the study considered the descending aorta aneurysmatic at a lower cut-off (≥35 mm) than is used in most studies.[27] Another study differentiated the segments in the thoracic aorta using a definition of aortic dilatation as a diameter above the 90th percentile of the aortic region according to reported reference values for age, sex, and body surface area, and aortic aneurysm defined as a diameter ≥50 mm for the ascending aorta and >40 mm for the descending aorta.[35] In this study, dilatation of the ascending aorta occurred in 28 of 144 patients (19.4%) and dilatation of the descending aorta occurred in 25 of 144 patients (17.4%). One patient met the definition of an aortic aneurysm, with a size of 52 mm in the ascending aorta.[35] As the development of descending aortic aneurysms seems to be rarely isolated and mainly coexists with ascending aortic aneurysms, we believe that ultrasonography of the ascending aorta is an excellent tool for monitoring the development or aortic aneurysms in patients with GCA, even if the descending part is not visualized.

Visual inspection of Figure 1 suggests greater variability in ultrasonography measurements compared with CT within the 35 to 40 mm CT diameter range, potentially indicating reduced agreement in this borderline dilation category. This may reflect technical limitations of ultrasonography in detecting subtle dilation or greater operator dependency in this intermediate range. In addition, ultrasonography may overestimate aortic diameter at higher values, particularly above 45 mm. This potential overestimation could impact clinical decision-making during follow-up, especially when approaching thresholds for intervention. In such cases, confirmatory imaging with CT may be warranted to ensure accurate assessment and avoid overtreatment.

In the prospective follow-up study, 20% of the patients, including patients of all GCA phenotypes, already had an aneurysm in the ascending aorta at baseline. It seems that aortic aneurysms at diagnosis could be part of a preclinical phase of the disease or present a genetic predisposition for aneurysm development in these patients, independent of the disease. Thus, we suggest that visualization of the aorta should be performed in all patients with GCA at diagnosis.

In the prospective cohort of 40 patients, the mean diameter of the ascending aorta increased by 2.6 mm during a follow-up time of 54 months. The aortic diameter in the general population has been shown to increase by 1 mm every 10 years.[36,37] Other studies of patients with GCA have also reported an increase in the diameter of the ascending and descending aorta more than expected.[8,10]

Several studies have suggested that LV involvement (LV-GCA or mixed-GCA) is a risk factor for developing aortic aneurysms. All patients in the prospective cohort of 40 who developed aneurysms had LV involvement, confirming previous studies that aneurysms occur in this group of patients.[4,5,11–13] Additionally, in the cross-sectional study cohort of 140 patients, a higher proportion of patients with LV involvement had an aneurysm; however, this difference was not statistically significant.

Our study demonstrated that male sex is a strong independent predictor for the development of ascending aortic aneurysm

in patients with GCA, with men exhibiting more than four-fold increased odds (OR 4.19, 95% CI 1.90–9.25, $P < 0.001$). This finding is well-aligned with previous evidence. Robson et al found that male sex was independently associated with twice the risk of aortic aneurysm (adjusted HR 2.10, 95% CI 1.38–3.19) in a large UK population-based cohort.[6] Furthermore, Mackie et al, through a systematic review and meta-analysis, identified male sex as a consistent risk factor for thoracic aortic aneurysm across multiple cohorts, highlighting a notably increased prevalence among men.[7] Collectively, these studies underscore the critical need for careful aortic surveillance in male patients diagnosed with GCA.

This study has limitations because of its real-life setting. The methodologic choices reflect real-world clinical practice but may affect measurement precision and intermodality comparability. Ultrasonography measurements were subject to operator-dependent variability, whereas CT data may have been influenced by differences in scan quality, phase of acquisition, contrast use, and measurement technique. Despite these limitations, the approach mirrors typical follow-up procedures for patients with GCA. Differences of 2 to 3 mm between examinations may occur without reflecting any real change in aortic size.[38] A limitation of the study is that one ultrasonographer (ACBH) was not blinded to a small number of CT scans performed before the ultrasonography examinations, which could have biased the ultrasonography measurements. Aneurysms in the abdominal aorta were not included. However, previous studies confirm that GCA mainly affects the thoracic aorta.[2,3,6,39] Other limitations include the small number of patients observed in the prospective study and the fact that imaging was not performed at predetermined intervals, with varying imaging modalities. Additionally, the use of a 40 mm cutoff for defining aortic aneurysms may be debated because aortic diameter may be influenced by individual factors such as age, sex, and body surface area.

The study's main strength is in its precise definition with a clear cut-off value for the aneurysm in the ascending aorta. It is also the first to compare CT and ultrasonography for measuring the diameter of this part of the aorta in patients with GCA. Another strength is the prospective follow-up of patients with GCA with a baseline image of the aorta at the time of GCA diagnosis.

In conclusion, ultrasonography yields results comparable with CT's in assessing the ascending aorta in patients with GCA. Aneurysms occurring alone in the descending part of the thoracic aorta are rare. Given ultrasonography's accessibility, cost-effectiveness, and lack of radiation, it could be a reasonable first-line tool for routine screening and monitoring of aortic aneurysms in this patient group. Patients with LV disease appear to have a higher risk for aneurysms and may need screening at regular intervals. Our data can inform further studies on the optimal interval, frequency, and modalities for aortic imaging in these patients.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Bull Haaversen 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. Jennette JC, Falk RJ, Bacon PA, et al. 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. Arthritis Rheum 2013;65(1):1–11.
2. Therkildsen P, de Thurah A, Nielsen BD, et al. Increased risk of thoracic aortic complications among patients with giant cell arteritis: a nationwide, population-based cohort study. Rheumatology (Oxford) 2022;61(7):2931–2941.
3. Evans JM, O'Fallon WM, Hunder GG. Increased incidence of aortic aneurysm and dissection in giant cell (temporal) arteritis. A population-based study. Ann Intern Med 1995;122(7):502–507.
4. de Boysson H, Aouba A. An updated review of cardiovascular events in giant cell arteritis. J Clin Med 2022;11(4):1005.
5. de Boysson H, Daumas A, Vautier M, et al. Large-vessel involvement and aortic dilation in giant-cell arteritis. A multicenter study of 549 patients. Autoimmun Rev 2018;17(4):391–398.
6. Robson JC, Kiran A, Maskell J, et al. The relative risk of aortic aneurysm in patients with giant cell arteritis compared with the general population of the UK. Ann Rheum Dis 2015;74(1):129–135.
7. Mackie SL, Hensor EM, Morgan AW, et al. Should I send my patient with previous giant cell arteritis for imaging of the thoracic aorta? A systematic literature review and meta-analysis. Ann Rheum Dis 2014;73(1):143–148.
8. García-Martínez A, Arguis P, Prieto-González S, et al. Prospective long term follow-up of a cohort of patients with giant cell arteritis screened for aortic structural damage (aneurysm or dilatation). Ann Rheum Dis 2014;73(10):1826–1832.
9. Kermani TA, Warrington KJ, Crowson CS, et al. Large-vessel involvement in giant cell arteritis: a population-based cohort study of the incidence-trends and prognosis. Ann Rheum Dis 2013;72(12):1989–1994.
10. Moreel L, Coudyzer W, Boeckxstaens L, et al. Association between vascular ^{18}F-fluorodeoxyglucose uptake at diagnosis and change in aortic dimensions in giant cell arteritis: a cohort study. Ann Intern Med 2023;176(10):1321–1329.
11. de Boysson H, Liozon E, Espitia O, et al. Different patterns and specific outcomes of large-vessel involvements in giant cell arteritis. J Autoimmun 2019;103:102283.
12. de Mornac D, Agard C, Hardouin JB, et al. Risk factors for symptomatic vascular events in giant cell arteritis: a study of 254 patients with large-vessel imaging at diagnosis. Ther Adv Musculoskelet Dis 2021;13:1759720X211006967.
13. Muratore F, Pazzola G, Pipitone N, et al. Large-vessel involvement in giant cell arteritis and polymyalgia rheumatica. Clin Exp Rheumatol 2014;32(3)(suppl 82):S106–S111.
14. Blockmans D, Coudyzer W, Vanderschueren S, et al. Relationship between fluorodeoxyglucose uptake in the large vessels and late aortic diameter in giant cell arteritis. Rheumatology (Oxford) 2008;47(8):1179–1184.

15. Haaversen AB, Brekke LK, Bakland G, et al. Norwegian society of rheumatology recommendations on diagnosis and treatment of patients with giant cell arteritis. Front Med (Lausanne) 2023;9: 1082604.
16. Dejaco C, Ramiro S, Duftner C, et al. EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice. Ann Rheum Dis 2018;77(5):636–643.
17. Dejaco C, Ramiro S, Bond M, et al. EULAR recommendations for the use of imaging in large vessel vasculitis in clinical practice: 2023 update. Ann Rheum Dis 2023;82:205.
18. Hellmich B, Agueda A, Monti S, et al. 2018 Update of the EULAR recommendations for the management of large vessel vasculitis. Ann Rheum Dis 2020;79(1):19–30.
19. Maz M, Chung SA, Abril A, et al. 2021 American College of Rheumatology/Vasculitis Foundation Guideline for the Management of Giant Cell Arteritis and Takayasu Arteritis. Arthritis Care Res (Hoboken) 2021;73(8):1071–1087.
20. Amsallem M, Ou P, Milleron O, et al. Comparative assessment of ascending aortic aneurysms in Marfan patients using ECG-gated computerized tomographic angiography versus trans-thoracic echocardiography. Int J Cardiol 2015;184:22–27.
21. Rodríguez-Palomares JF, Teixidó-Tura G, Galuppo V, et al. Multimodality assessment of ascending aortic diameters: comparison of different measurement methods. J Am Soc Echocardiogr 2016;29(9): 819–826.e4.
22. Dejaco C, Duftner C, Buttgereit F, et al. The spectrum of giant cell arteritis and polymyalgia rheumatica: revisiting the concept of the disease. Rheumatology (Oxford) 2017;56(4):506–515.
23. Hunder GG, Bloch DA, Michel BA, et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis Rheum 1990;33(8):1122–1128.
24. Terslev L, Diamantopoulos AP, Døhn UM, et al. Settings and artefacts relevant for Doppler ultrasound in large vessel vasculitis. Arthritis Res Ther 2017;19(1):167.
25. Patel HN, Miyoshi T, Addetia K, et al. Normal values of aortic root size according to age, sex, and race: results of the World Alliance of Societies of Echocardiography Study. J Am Soc Echocardiogr 2021; 34(10):1113–1123.e5.
26. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2015;16(3):233–270.
27. Hiratzka LF, Bakris GL, Beckman JA, et al; Society for Vascular Medicine 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with Thoracic Aortic Disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, American Association for Thoracic Surgery, American College of Radiology, American Stroke Association, Society of Cardiovascular Anesthesiologists, Society for Cardiovascular Angiography and Interventions, Society of Interventional Radiology, Society of Thoracic Surgeons, and Society for Vascular Medicine. Circulation 2010; 121(13):e266–e369.
28. Mazzolai L, Teixido-Tura G, Lanzi S, et al; ESC Scientific Document Group 2024 ESC Guidelines for the management of peripheral arterial and aortic diseases. Eur Heart J 2024;45(36):3538–3700.
29. Borgbjerg J, Christensen HS, Al-Mashhadi R, et al. Ultra-low-dose non-contrast CT and CT angiography can be used interchangeably for assessing maximal abdominal aortic diameter. Acta Radiol Open 2022;11(10):20584601221132461.
30. Wanhainen A, Van Herzeele I, Bastos Goncalves F, et al; ESVS Guidelines Committee; Document Reviewers. Editor's Choice -- European Society for Vascular Surgery (ESVS) 2024 Clinical Practice Guidelines on the Management of Abdominal Aorto-Iliac Artery Aneurysms. Eur J Vasc Endovasc Surg 2024;67(2):192–331.
31. Nuenninghoff DM, Hunder GG, Christianson TJH, et al. Incidence and predictors of large-artery complication (aortic aneurysm, aortic dissection, and/or large-artery stenosis) in patients with giant cell arteritis: a population-based study over 50 years. Arthritis Rheum 2003; 48(12):3522–3531.
32. Gonzalez-Gay MA, Garcia-Porrua C, Piñeiro A, et al. Aortic aneurysm and dissection in patients with biopsy-proven giant cell arteritis from northwestern Spain: a population-based study. Medicine (Baltimore) 2004;83(6):335–341.
33. García-Martínez A, Hernández-Rodríguez J, Arguis P, et al. Development of aortic aneurysm/dilatation during the followup of patients with giant cell arteritis: a cross-sectional screening of fifty-four prospectively followed patients. Arthritis Rheum 2008;59(3):422–430.
34. Gallou S, Agard C, Dumont A, et al. Evolution and outcomes of aortic dilations in giant cell arteritis. Eur J Intern Med 2024;129:71–77.
35. Jud P, Verheyen N, Dejaco C, et al. Prevalence and prognostic factors for aortic dilatation in giant cell arteritis - a longitudinal study. Semin Arthritis Rheum 2021;51(4):911–918.
36. Rogers IS, Massaro JM, Truong QA, et al. Distribution, determinants, and normal reference values of thoracic and abdominal aortic diameters by computed tomography (from the Framingham Heart Study). Am J Cardiol 2013;111(10):1510–1516.
37. Agmon Y, Khandheria BK, Meissner I, et al. Is aortic dilatation an atherosclerosis-related process? Clinical, laboratory, and transesophageal echocardiographic correlates of thoracic aortic dimensions in the population with implications for thoracic aortic aneurysm formation. J Am Coll Cardiol 2003;42(6):1076–1083.
38. Elefteriades JA, Mukherjee SK, Mojibian H. Discrepancies in measurement of the thoracic aorta: JACC review topic of the week. J Am Coll Cardiol 2020;76(2):201–217.
39. ES. M. Cardiovascular outcomes and in-hospital mortality in giant cell arteritis. Case Western Reserve University; 2008.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 537–546
DOI 10.1002/acr.25653

American College of Rheumatology
Empowering Rheumatology Professionals

Impact of Positive Lifestyle Behaviors on Direct Health Care Cost Savings for Low Back Pain

Ye Tian,[1] Katharine E. Roberts,[1,2] Michelle Hall,[2] Paula R. Beckenkamp,[1] Angelo Sabag,[1] Karoline Moe,[3] Ana Paula Carvalho-e-Silva,[1] Emily J. Callander,[4] and Paulo H. Ferreira[1]

Objective. This study aimed to investigate the relationship between a previously purpose-developed lifestyle behavior scale and health care cost savings related to low back pain (LBP).

Methods. This longitudinal study used data from the Australian Twin Back (AUTBACK) study. LBP and lifestyle behavior measures were collected at baseline. Physical activity (PA) was objectively quantified via an accelerometer. A lifestyle behavior scale was created using variables of body mass index, PA, smoking status, and sleep quality. Weekly health care use for LBP was collected over one year. Health care costs were calculated by aggregating expenses for health care visits and medications, encompassing personal and Australia Medicare costs, and analyzed by two-part models.

Results. Individuals with lower lifestyle behavior scores, women, and those with a baseline episode of LBP were more likely to incur health care use costs (n = 307). A total of 2.6% of participants accounted for over 56% of the total expenditures. A one-point improvement in the lifestyle behavior scale was significantly associated with 23% decrease in overall health care costs for LBP (95% confidence interval [CI] 7%–36%; $P = 0.006$), 25% decrease in medication costs for LBP (95% CI 13%–35%; $P < 0.001$), and 27% decrease in health care visit costs for LBP (95% CI 14%–39%; $P < 0.001$). The predicted difference in yearly health care use costs between individuals with the lowest and highest lifestyle scores was AU$873.

Conclusion. This study demonstrated the association between greater adherence to positive lifestyle behaviors and reduced health care costs related to LBP. Interventions aimed at improving lifestyle behaviors could yield substantial cost savings for individuals and the health care system, mitigating the burden of LBP.

INTRODUCTION

Low back pain (LBP) affects 619 million people globally and is the leading cause of years lived with disability (YLDs), accounting for a total of 69 million YLDs worldwide.[1] Modifiable risk factors account for approximately two-fifths of LBP-related YLDs, with 12.5% of YLDs attributed to smoking, and 11.5% attributed to high body mass index (BMI).[1] LBP also constitutes a significant economic burden on the individual and health care system. The average annual direct cost estimates, including costs for goods and health care services due to LBP, in high-income countries range from US$4,671 to US$10,430 per patient with LBP.[2] The Australian Institute of Health and Welfare reported that, in 2020 to 2021, Australians spent an estimated AU$3.4 billion on managing back problems,[3] encompassing direct and indirect costs. Direct costs include expenses related to services and goods used in the diagnosis, management, and rehabilitation of LBP, whereas indirect costs represent the economic burden resulting from reduced work productivity and absenteeism.[2]

Lifestyle behaviors including smoking, alcohol consumption, and poor sleep have been found to be associated with higher health care use for LBP.[4] A randomized controlled trial conducted

The funder had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Supported by the National Health and Medical Research Council. The Australian Twin Back study was supported by the Sydney University National Health and Medical Research Council (NHMRC; bridging grant GNT1086738). Dr Hall's work was supported by an NHMRC Investigator Grant (1172928). Twins Research Australia was supported by the NHMRC through a Centre of Research Excellence Grant (GNT1079102), which was administered by the University of Melbourne.

[1]School of Health Science, Faculty of Medicine and Health, University of Sydney, Sydney, New South Wales, Australia; [2]Sydney Musculoskeletal Health, The Kolling Institute, University of Sydney, Sydney, New South Wales, Australia; [3]Norwegian University of Science and Technology, Trondheim, Norway; [4]School of Public Health, University of Technology Sydney, Sydney, New South Wales, 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.25653).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25653.

Address correspondence via email to Ye Tian, BAppSc(Phty)(Hons), at tiara.tian@sydney.edu.au.

Submitted for publication April 9, 2025; accepted in revised form September 2, 2025.

SIGNIFICANCE & INNOVATIONS

- The study demonstrates an association between greater adherence to positive lifestyle behaviors and reduced yearly direct health care costs due to low back pain (LBP). The potential savings can reach almost AU$900 in yearly overall health care use costs when individuals with lowest and greatest adherence to positive lifestyles are compared.
- A small subset (2.6%) of participants drove more than 56% of the total health care costs. This underscores the potential for targeted interventions focused on this subset of populations to yield substantial cost savings in LBP management.
- The study findings highlight the potential for self-management and clinical and policy strategies promoting positive lifestyle behaviors to yield substantial cost savings in LBP management.

in Australia found that interventions promoting healthy lifestyle behaviors, including weight loss and increased physical activity (PA), were associated with lower health care costs when compared with usual care (AU$708, 95% confidence interval [CI] AU $581–AU$850).[5] A growing body of evidence supports the reallocation of health care resources from treatment of chronic conditions toward positive health initiatives or lifestyle interventions, with approaches such as health education and lifestyle changes resulting in improved health outcomes in chronic disease including cardiovascular disease, diabetes, and cancer.[6] For instance, changes in lifestyles, including lowering energy intake via a healthy diet, and increasing energy expenditure via engagement in moderate intensity walking for 30 minutes five days per week, prevented the onset of type 2 diabetes by 58% for Americans at high risk.[7] Interestingly, clusters of healthier lifestyles have also been found to be associated with lower levels of psychological distress, higher quality of life, and better self-rated health.[8]

The novelty and use of a positive lifestyle behavior scale (PLBS) in LBP has been established in our previous study, with results revealing that increases in PLBS scores are associated with less health care usage for LBP over 12 months.[9] Apart from the benefits in clinical outcomes, evidence linking healthy lifestyle factors to reduced health care costs for LBP could also encourage governments and policymakers to shift their focus toward positive health interventions, therefore leading toward long-term cost savings and improved health outcomes at the individual and population levels. In the present study, we aimed to determine whether engagement in PA, sleep quality, smoking status, and BMI levels, measured by the PLBS, was associated with participants' yearly direct health care use cost, including out-of-pocket expenses and items covered by Australia Medicare, due to LBP.

MATERIALS AND METHODS

Study design. Data for this study derived from the Australian Twin Back (AUTBACK) study, a longitudinal cohort study which aimed to evaluate the relationship between PA and LBP outcomes.[10] The study was approved by the Twin Research Australia and the University of Sydney Human Ethics Committee (project 2015/407).

Participants. Between 2015 and 2020, twins aged >18 years old (n = 401) were recruited across urban, remote, and rural Australia via Twins Research Australia. Eligible participants were twins aged >18 years old, with internet access via phone or computer and active email accounts. Individuals with self-reported serious spinal pathology (eg, inflammatory, metastatic, or infectious disease of the spine), pregnant women, or those with a recent history of spinal surgery (≤12 months) were excluded. In the AUTBACK study, data were collected at baseline (ie, study enrollment) and six-month follow-up via self-reported questionnaires (eg, demographic and anthropometric information, lifestyle behaviors, and LBP history) and accelerometer devices (eg, objective PA). Self-reported measures including LBP symptoms (eg, pain severity, activity limitation), health care usen (eg, physiotherapist, surgery), self-management strategies (eg, heat pack, light exercise) and medication (eg, over-the-counter medications, opioids) were collected weekly over 12 months.[10] Details of the weekly electronic questionnaire are available in Supplementary A.1. In the current study, only participants who reported a lifetime prevalence of LBP (ie, LBP at any point in life) (Supplementary A.2) and provided both baseline data on BMI, smoking status, sleep variables, and PA levels and weekly data on health care utilization due to LBP were included. All participants provided informed consent before data collection.

Assessment of exposures. A lifestyle behavior scale (PLBS) based on lifestyle factors including BMI, smoking status, PA, and sleep quality, which has been shown to predict health care usage related to LBP, was used to measure the level of healthiness in lifestyles.[9] These four lifestyle factors were categorized as optimal (two points), intermediate (one point), or poor (zero points) based on prespecified thresholds to align with World Health Organization (WHO) recommendations and IPD-Work (individual-participant data meta-analysis in working populations) Consortium articles.[11–13] Equal weighting was used to avoid overemphasizing any single lifestyle factor. The individual scores were then summed into a composite scale ranging from zero (least adherent to positive lifestyle behaviors) to eight (most adherent to positive lifestyle behaviors). Details of the scoring system used for computing PLBS are available in Supplementary A.3.

BMI was calculated using weight and height measurements self-reported at baseline. BMI between 18.5 (inclusive) and 25.0 kg/m^2, indicating normal weight, was categorized as optimal;

BMI between 25.0 (inclusive) and 30.0 kg/m^2 was categorized as intermediate; BMI lower than 18.5 kg/m^2 was categorized as underweight; and BMI of 30 kg/m^2 and higher, indicating obesity, was categorized as poor.[11] Baseline data on smoking status were collected, with nonsmokers categorized as optimal, exsmokers or occasional smokers categorized as intermediate, and current smokers categorized as poor. Participants wore an Actigraph accelerometer (GT1M/GT3X model) on their right hip for seven consecutive days during their waking hours. The methodology for processing PA variables can be found in Supplementary A.4. According to the WHO recommendations for PA,[12] optimal levels were defined as either ≥150 minutes of moderate PA per week, ≥75 minutes of vigorous PA per week, or a combination of both. Moderate PA of 60 to 150 minutes or vigorous PA of 20 to 70 minutes per week were categorized as intermediate, and moderate PA of <60 minutes or vigorous PA of <20 minutes per week were categorized as poor.[13,14] The Pittsburgh Sleep Quality Index (PSQI), a standard clinical instrument with scores ranging from 0 to 21, was used to assess sleep quality.[15] The sleep quality and daytime dysfunction subscales, each constituting one-seventh of the PSQI, were used for the PLBS because they are thought to correlate with LBP prognosis. A sleep quality score of zero and daytime dysfunction score of zero or one were categorized as optimal, sleep quality score of one and daytime dysfunction score of one or two were categorized as intermediate, and sleep quality of two or three and daytime dysfunction of two or three were categorized as poor.[9]

Covariates. The following variables were considered as possible covariates, given their potential to influence the prognosis of LBP and health care use due to LBP in previous studies: sex, age, depression, anxiety, stress, and a history of LBP in the previous four weeks.[16–18] Depression, anxiety, and stress were assessed with 21-item Depression Anxiety Stress Scales,[19] Data on all covariates were collected via the baseline questionnaire.

Assessment of health care use and health care costs due to LBP. Data on the number of LBP flare-ups and care-seeking behaviors associated with LBP were collected weekly, over 12 months, via electronic questionnaires (Supplementary A.1) sent via SMS or email. Each week with reported LBP was defined as a distinct LBP episode, for which the corresponding health care use cost was ascertained. Overall health care use cost due to LBP was calculated by aggregating the total expenses incurred from health care visits and medication usage for LBP over a 12-month period. The health care costs encompass both out-of-pocket expenses and items covered by Medicare, the government-funded health care insurance program in Australia.

A hierarchical decision process was employed when calculating health care visit and medication costs. Health care services were valued using unit service prices of the Australian Medicare Benefit Schedule (MBS).[20] Medication usage were valued by (1) applying the conservative daily dosages based on standard Australian medication guideline[21] to the reported days, (2) annualizing quantities and rounding to whole package sizes, and (3) applying unit prices of the Australian Pharmaceutical Benefits Scheme (PBS).[21] If the MBS or PBS unit prices were unavailable, prices were obtained by computing the average prices from the top five to 10 relevant private clinic services or pharmacy items in two major Australian states, Victoria and New South Wales. All health care costs are reported in Australian dollars (AU$1 = US $0.656 or €0.606) and reflected the costs of goods and services in the 2023 to 2024 financial year.

Statistical analysis. Descriptive statistics were conducted and presented as medians and interquartile ranges (IQR) due to non-normal distribution of the data. For variables presenting frequently zero values, mean and SD was reported. Based on previous research, due to the large fraction of observations with zero expenditures, a two-part model was used to calculate adjusted associations between the PLBS and the primary outcome. In the two-part model, logistic regression (logit) was used to model the probability that a person has any health care expenditures using the full sample.[22] Then, we modeled the association between PLBS and the primary outcome with a generalized linear model (GLM), interoperating a γ-distribution and a log-link function, on the subset of people who had any expenditures.[22] Specifically, to quantify the economic impact of lifestyle behaviors, we calculated the adjusted mean annual LBP-related health care use costs across PLBS levels (0–8) using the "margins" command following the two-part model. The margins command computed unconditional marginal predictions by evaluating costs at each integer PLBS value while preserving observed covariate distributions, thereby accounting for cohort heterogeneity. Adjusted models accounted for the potential confounding effects of sex, age, depression, anxiety, stress, and a history of LBP in the previous four weeks at baseline. Adjusted odds ratios (ORs), cost ratios (CRs), and 95% CIs were used to quantify the associations between the exposure variables and the study outcome. To account for the nonindependence of data from complete twin pairs, robust standard errors were calculated using the "vce cluster" command, which clusters on the identifier distinguishing twins within a pair. Statistical significance was set at $P < 0.05$. All analyses were performed using Stata (version 18).[23]

Data sharing. The dataset generated during and/or analyzed during the current study are included in the article or uploaded as online supplementary information. Deidentified individual participant data will be made available, including data dictionaries, for approved data sharing requests. Individual participant data will be shared that underlie the results reported in this article, after deidentification and normalization of information (text, tables, figures, and appendixes). Study protocol has been published and is publicly available. Anonymized data will

be available beginning three months after and ending 10 years after publication of this article to researchers who provide a completed Data Sharing Agreement that describes a methodologically sound proposal for the purpose of the approved proposal.

RESULTS

Overall, 16,690 weekly questionnaires assessing LBP status and health care use costs due to LBP were completed across 401 eligible participants, corresponding to a response rate of 80% (16,690 questionnaires received out of 20,852 [401 participants × 52 weeks]). A total of 59 participants who did not report a lifetime prevalence of LBP were excluded. Only two participants were excluded due to an unavailable weekly dataset, resulting in minimal impact on study generalizability. A total of 33 participants were excluded from the analyses due to unavailable Actigraph data (Figure 1). No multiple imputation was conducted in the analyses.

Baseline characteristics. A total of 307 participants were included in the current study. Over the 12-month study period, participants collectively reported 3,094 weeks or 11,334 days with LBP. On average, each participant experienced LBP for 10.1 days (SD 12.7). The baseline characteristics of the sample are presented in Table 1. The median age was 56 years (IQR 45.0–62.3), and most participants (74%) were women. A total of 48% of the study sample reported a recent episode of LBP in the four weeks before baseline completion; 34% (47 of 139) of those participants reported experiencing LBP for six weeks or less, 4% (6 of 139) reported experiencing LBP for six weeks to three months, and 62% (86 of 139) reported experiencing LBP for three months or more. The median BMI was 24.6 kg/m^2 (IQR 22.1–28.1), and 52% of participants had BMI between 18.5 and 24.9 kg/m^2 and were categorized as having optimal BMI. A total of 55% of participants were categorized as having optimal PA (ie, moderate PA of ≥150 minutes and/or vigorous PA of ≥75 minutes per week). Although only 22% of participants were categorized as having optimal sleep status (ie, great sleep quality and minimal daytime dysfunction), 82% of participants were nonsmokers and categorized as having optimal smoking status. The median PLBS was six (IQR 5–7), with the majority of participants (71%) scoring five, six, or seven (Figure 2). A total of 8% of the participants (n = 24) scored the maximum PLBS of eight, no

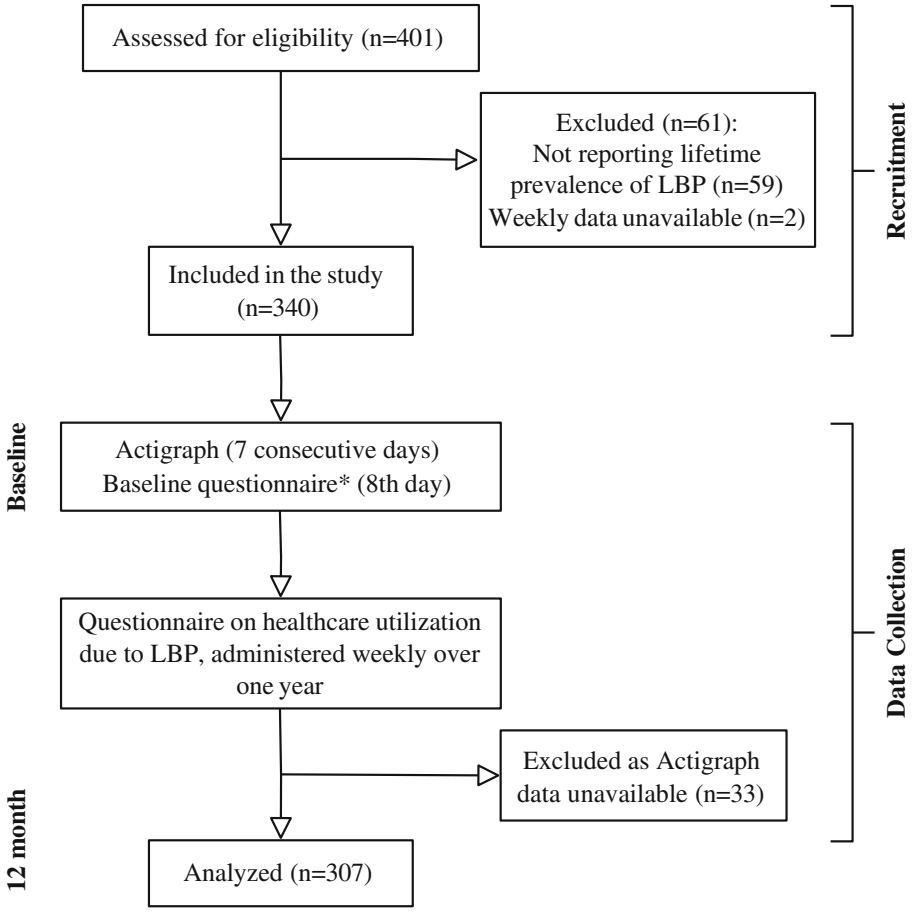

Figure 1. Study flowchart. *Exposure variables were collected via the baseline questionnaire, including body mass index, smoking status, sleep quality, and physical activity. Covariates were collected via the baseline questionnaire, including sex, age, depression, anxiety, stress, and a history of LBP in the previous four weeks. LBP, low back pain.

Table 1. Baseline characteristics of the participants*

Characteristic	n	Values
Age, median (IQR)	307	55.8 (45.0–62.3)
Female, %	307	74[a]
Body mass index, median (IQR), kg/m^2	307	24.6 (22.1–28.1)
Zygosity (monozygotic), %	307	66[b]
Recent episode of LBP (yes)[c], %	307	48[d]
LBP duration[e], %		
≤6 weeks	47	34
6 weeks to 3 months	6	4
≥3 months	86	62
Depression (0–42)[f], median (IQR)	307	2 (0–4)
Anxiety (0–42)[f], median (IQR)	307	2 (0–4)
Stress (0–42)[f], median (IQR)	307	6 (2–12)
Positive lifestyle variables, %		
BMI categories		
Optimal	159	52
Intermediate	98	32
Poor	50	16
Smoking status categories, %		
Optimal	251	82
Intermediate	50	16
Poor	6	2
Sleep categories, %		
Optimal	68	22
Intermediate	172	56
Poor	67	22
Physical activity categories[g], %		
Optimal	170	55
Intermediate	97	32
Poor	40	13
Positive lifestyle behavior scale, median (IQR)	307	6 (5–7)
1, %	1	0.33
2, %	9	2.93
3, %	19	6.19
4, %	36	11.73
5, %	77	25.08
6, %	71	23.13
7, %	70	22.80
8, %	24	7.82

* BMI, body mass index; IQR, interquartile range; LBP, low back pain.
[a] n = 226.
[b] n = 204.
[c] Recent episode of LBP is defined as experiencing low back pain ≤4 weeks before completion of baseline assessment.
[d] n = 146.
[e] Data on LBP duration were collected when individuals reported recent episode of LBP.
[f] 21-Item Depression Anxiety Stress Scale; each of the three domains range from 0–42, with higher scores representing higher levels of each domain.
[g] Device-based measures; values are minutes/week, assessed with an accelerometer.

participant presented the minimum score of zero, and only one participant (0.3%) scored one.

The impact of PLBS on health care costs for LBP. The median overall health care cost due to LBP over 12 months was AU$15.78 (IQR 0–110.35), with the highest median cost of AU $41.39 (IQR 15.78–528.18) when PLBS was scored at three. Those who scored one, seven, or eight had an overall health care cost of AU$0 (Table 2). Out of 307 participants, 173 (56%) had an overall health care cost over AU$0. The primary health care visit cost driver was physiotherapy visits (mean AU$67.48, SD AU$462.03), and the primary medication cost drivers were usage of weak opioids (mean AU$18.19, SD AU $117.75) and nonopioids (mean AU$11.10, SD AU$22.17). Only 2.6% of participants (n = 8) had an overall health care cost higher than AU$1,000, and 1.3% of the participants (n = 4) had an overall health care cost exceeding AU$4,000. Details of the distribution and breakdown of health care use costs are available in Supplementary Figure 1 and Supplementary Tables 1 and 2.

Residual plot (Supplementary Figure 2) indicated linearity across 96.7% of the PLBS distribution (scores ≥2), with minor deviations only at the lower extreme (scores ≤2; n = 10, 3.3% of observations). Given both the limited representation of PLBS of two and lower and the absence of meaningful deviation from linearity above this threshold, we kept the PLBS variable as a continuous linear variable in all analyses.

In the full sample (n = 307), the association between PLBS and the probability of having health care expenditures larger than AU$0 is presented in Table 3 (logit column). Higher scores on the PLBS were significantly associated with a lower probability of incurring overall health care costs (OR 78%, 95% CI 65%–94%; $P = 0.008$) and a decreased probability of incurring medication costs (OR 79%, 95% CI 66%–94%; $P = 0.008$). This indicates that an increase in PLBS was associated with a decreased chance of incurring any expenditures (>AU$0) on LBP.

In the subset of people who had any expenditures due to LBP (n = 173), the association between PLBS and overall health care costs, medication costs, and health care visit costs due to LBP is presented in Table 3 (GLM column). Among those who spent more than AU$0 due to LBP, a one-point improvement in the PLBS was significantly associated with a 23% decrease in overall health care costs (CR 0.77, 95% CI 0.64–0.93; $P = 0.006$), a 25% decrease in medication costs (CR 0.75, 95% CI 0.65–0.87; $P < 0.001$), and a 27% decrease in health care visit costs due to LBP (CR 0.73%, 95% CI 0.61–0.86; $P < 0.001$). Being a man was significantly associated with a 76% lower overall health care costs (CR 0.24, 95% CI 0.11–0.54; $P < 0.001$) compared with being a women (Supplementary Table 3). Having a baseline episode of LBP was significantly associated with a 208% increase in overall health care costs (CR 3.08, 95% CI 1.87–5.07; $P < 0.001$) (Supplementary Table 3).

Predicted health care use costs. For participants with the median (six) PLBS, the predicted overall health care costs over 12 months was AU$147.30 (95% CI 85.00–209.60; $P < 0.001$). The predicted overall health care costs were highest (AU $946.49; 95% CI −133.66 to 2026.65; $P = 0.086$) for those with

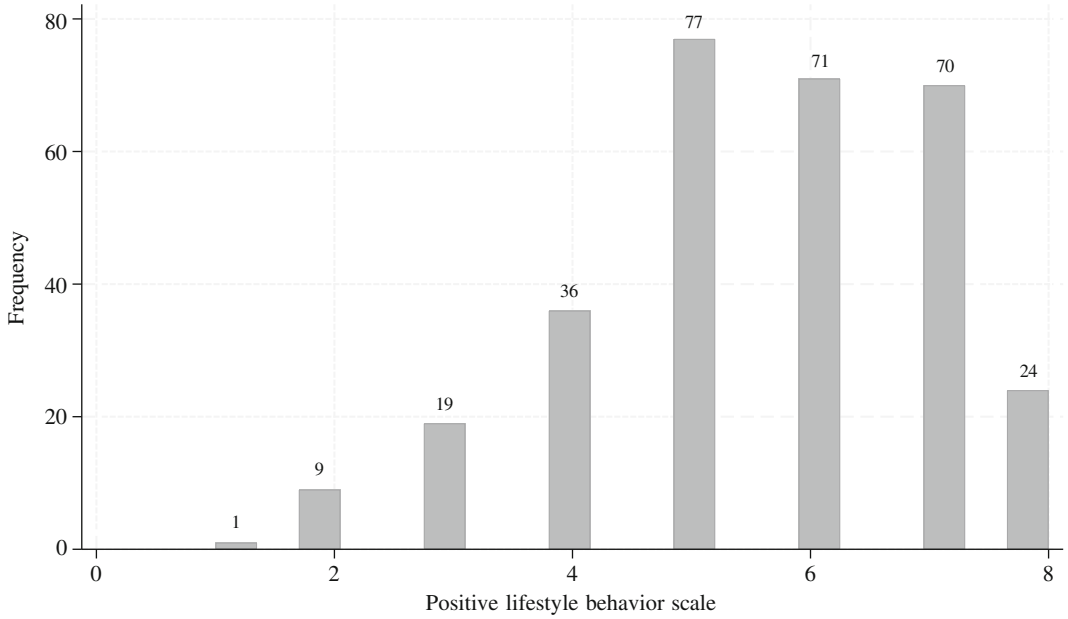

Figure 2. Histogram showing the distribution of positive lifestyle behavior scale.

Table 2. Descriptive analysis of positive lifestyle behavior scale and overall health care use costs due to low back pain*

Positive lifestyle behavior scale[a]	n	Cost[b], median (IQR)	Cost, mean (SD)
Total	307	15.78 (0–110.35)	212.36 (891.37)
0	0	N/A	N/A
1	1	0 (0–0)	0 (N/A)
2	9	15.78 (0–15.78)	15.78 (24.95)
3	19	41.39 (15.78–528.18)	1,100.41 (2,483.28)
4	36	15.78 (0–119.24)	380.36 (1,606.31)
5	77	15.78 (0–102.51)	152.82 (442.49)
6	71	15.78 (0–252.25)	161.45 (279.60)
7	70	0 (0–82.90)	86.88 (186.25)
8	24	0 (0–15.78)	47.49 (142.25)

* IQR, interquartile range; N/A, not applicable.
[a] Total lifestyle behavior scale ranges from 0 to 8 (where 0 represents the lowest positive lifestyle behavior scale, and 8 represents the highest positive lifestyle behavior scale).
[b] Overall health care use cost is defined as the total expenses incurred from health care practitioner visits and medication usage over the 12-month study period.

Table 3. The association between the positive lifestyle behavior scale, covariates and overall costs, medication costs, and health care visit costs due to low back pain*

	Logit (n = 307)		GLM (n = 173)	
Exposure	OR (95% CI)	*P* value	CR (95% CI)	*P* value
Positive lifestyle behavior scale[a]				
Overall health care costs[b]	0.783 (0.654–0.938)	0.008	0.771 (0.641–0.927)	0.006
Medication costs[b]	0.789 (0.662–0.940)	0.008	0.754 (0.650–0.874)	<0.001
Health care visit costs[b]	0.943 (0.788–1.129)	0.52	0.727 (0.614–0.860)	<0.001

* CI, confidence interval; CR, cost ratio; GLM, generalized linear model; OR, odds ratio.
[a] Positive lifestyle behavior scale was treated as a continuous variable in these analyses.
[b] All analyses were adjusted for sex, age, depression, anxiety, stress, and recent episode of low back pain at baseline.

the minimum (zero) PLBS. On the other hand, those who scored eight on the PLBS had the lowest predicted overall health care costs of AU$73.34 (95% CI 28.81–117.87; $P = 0.001$) (Figure 3; Supplementary Table 4). The predicted difference in overall health care costs between individuals with the lowest and highest lifestyle scores was AU$873 (Supplementary Table 4). For participants with the median (six) PLBS, the predicted medication costs were AU$18.70 (95% CI 5.98–31.42; $P = 0.004$), and the predicted health care visit costs were AU$125.39 (95% CI 67.12–183.67; $P < 0.001$). For participants with the minimum (zero) PLBS, the predicted medication costs (AU $149.42; 95% CI −91.83 to 390.67; $P = 0.225$) and predicted health care visit costs (AU$1,028.44; 95% CI −250.09 to 2,306.98; $P = 0.115$) were highest (Figure 3; Supplementary Table 4).

DISCUSSION

In the present study, we observed positive associations between people's positive lifestyle behaviors, assessed by the PLBS, and a reduction in their annual health care use costs—such that optimal lifestyle behaviors were associated with significantly lower overall health care, medication, and health care visit costs for LBP. These costs were primarily derived from savings associated with physiotherapy sessions, as well as medication usage, such as weak opioids and nonopioids. As individuals adopt more positive lifestyle behaviors, the predicted annual health care costs associated with LBP demonstrated a corresponding decrease. These findings suggest that for individuals with a lifetime history of LBP, adherence to positive lifestyle behaviors may lead to significant cost savings in annual health care expenditures related to LBP. For governments and policymakers, these findings underscore the importance of increasing access to positive lifestyle interventions and promoting positive health initiatives.

In line with existing evidence,[18,24] 56% of the study participants, with a lifetime history of LBP, chose to seek care and pay for LBP treatment. Despite the burden of LBP, approximately 56% to 60% of people suffering from LBP seek health care.[18,24] This low health care–seeking rate is potentially due to the self-limiting nature of acute episodes of LBP, which often resolves with self-care strategies.[24] In the current study, yearly health care cost savings incurred from health care professional visits and medication usage were strongly associated with positive lifestyle behaviors, including healthy BMI, minimal smoking, increased PA, and optimal sleep quality. A one-point improvement in people's lifestyle behavior score was significantly associated with a 23% decrease in their overall health care costs, a 25% decrease in medication costs, and a 27% decrease in health care visit costs due to LBP. These findings are in line with a previous interventional study showing that weight loss and optimal PA adherence resulted in AU$708 (95% CI 581–850) lower annual health care costs in adults with chronic LBP when compared with usual care.[5] In our study, men, compared with women, used 76% lower overall health care costs. This finding is consistent with a previous meta-analysis, in which pooled results showed that sex is an important determinant in the care-seeking behavior.[18] Sex differences in costs may reflect both biologic factors (eg, higher LBP prevalence in women) and sociocultural influences (eg, caregiver roles traditionally being adopted by women). According to the 2023 Global Burden of Disease Report, the global LBP prevalence rates were higher among women compared with men across all age groups.[1] Compared with men, women have worse self-perceived health, which could account for their greater use of health care services, such as primary care visits and use of diagnostic procedures.[25] Moreover, women frequently assume the caregiver roles of the family and, therefore, are more heightened in health awareness and active in care-seeking behaviors.[26] These sex differences underscore the need to adopt sex sensitivity into health care practice by raising awareness in public health, enhancing clinical understanding, and developing skills among health professionals[27] to ensure equitable care delivery while addressing potentially modifiable cost drivers.

Only a small proportion of the sample (2.6%) had an overall health care cost higher than AU$1,000, and 1.3% had an overall health care cost higher than AU$4,000. Individuals who spent over AU$1,000 annually on LBP care accounted for >56% of the total expenditure within the study sample. These findings suggest that a relatively small subset of the population drives most health care expenditure related to LBP. Participants that drove the higher costs (ie, that spent over AU$4,000) were women, had relatively low PLBS scores (median 3, IQR:3–3.5), most commonly saw the physiotherapist of all health care practitioners, most commonly used weak opioids of all medications for LBP, and all had three months or more of LBP duration and higher stress level (Supplementary Table 2). A previous study and a systematic review showed that, rather than socioeconomic status, factors that increased care-seeking behaviors for LBP were older age, female sex, higher grades of pain and disability, and fear of the impact of pain on participation with usual social roles.[18,28] Policymakers could target those with higher risks for health care costs by promoting specific health care plans, including community-based exercise and disability benefit and support programs.

Although people who adopt healthier lifestyle behaviors might not spend less time suffering from activity limiting LBP,[9] they are likely to use less health care and increase their cost savings related to LBP. Simple lifestyle changes, such as walking in the community, healthier eating, quitting smoking, or improving sleeping habits, could help individuals to significantly reduce the financial burden of LBP management. For individuals with LBP and clinicians who manage LBP, incorporating lifestyle programs could empower patients to adopt diverse self-care strategies, improve LBP health outcomes, and increase cost savings.

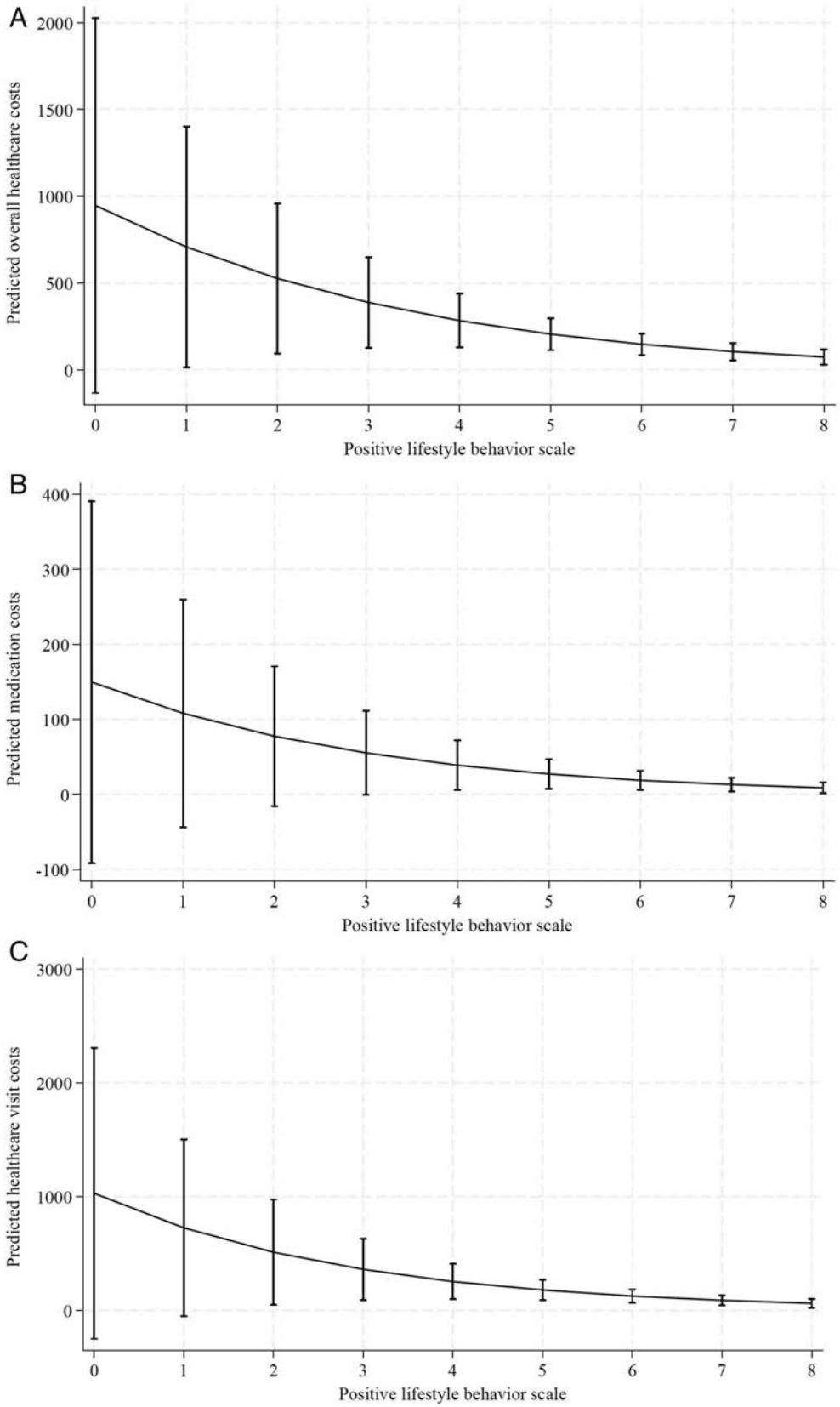

Figure 3. Margins plot showing the predicted health care use costs with confidence interval at each level of positive lifestyle behavior scale for: (A) overall health care costs, (B) medication costs, and (C) health care visit costs due to low back pain.

Components of lifestyle programs identified for addressing LBP include healthy diet, exercise and leisure-time PA, sleep hygiene, stress resilience (eg, mindfulness, diaphragmatic breathing), reduced substance use (eg, nonsmoking status, moderate alcohol consumption), social connection (eg, community activities, exercise buddy), and self-management strategies (eg, self-massage, flare-up management).[29] The effectiveness of these interventions can vary depending on factors like individual preferences, cultural contexts, and socioeconomic conditions. Therefore, a patient-centered approach is necessary to tailor lifestyle interventions to specific populations and maximize the positive impact over time.

For policymakers, prioritizing healthy lifestyle interventions via advocating positive lifestyle behaviors plays a crucial role in promoting overall health, enhancing quality of life by preventing chronic disease, reducing disability level, and alleviating health care economic burdens at both an individual and societal level. For instance, the Australian National Tobacco Strategy 2023 to 2030 is in place to reduce the use of both tobacco and e-cigarettes.[30] Many Australians consume excessive energy through unhealthy diets.[31] The WHO advises that reasonable price policies on sugar-sweetened beverages and subsidies for fresh fruits or vegetables are effective in improving the affordability and consumption of healthier food products, hence improving BMI-related health outcomes.[32] In Newfoundland and Labrador, Canada, the Physical Activity Tax Credit offers a refundable tax credit of up to CA$2,000 per household, which acts as an incentive for families as they access sport and leisure-time PA.[33] Although a previous study has shown that national health campaigns have a beneficial impact on the adoption of various positive lifestyle behaviors,[34] policymakers should also acknowledge, and target, potential barriers in the implementation of healthy lifestyle interventions, for example, by improving accessibility and affordability to positive lifestyle services (eg, Healthdirect, Alcohol and Drug Information Hotline, community-based exercise programs) and raising public awareness via empowering patient education. People are more likely to make healthy lifestyle choices when they are enabled and empowered with the necessary knowledge, motivation, and support.

The AUTBACK study included participants across all regions of Australia, and the response rate was high (80%), providing a large nationwide database. The longitudinal nature of the study allowed us to collect repeated weekly measures of health care use over a 12-month period. Additionally, device-based measures of PA were used, which demonstrated acceptable reliability and validity for assessing PA levels.[35] Whereas earlier work established that healthier lifestyles reduced health care usage frequency,[9] our study moved beyond evenly weighted health care usage counts (eg, one visit to physiotherapist or one opioid consumption) and uniquely quantified the substantial cost saving, accounting for varied service and good prices, and identified high-cost subgroups for optimized resource allocation. For example, our analysis revealed that reduced health care use translated to a maximum predicted cost saving of AU$873.15 per patient annually when accounting for service-specific cost variations. Furthermore, our identification of the distinct cost-profile subgroups (ie annual health care utilization cost larger than AU$4,000) reported in Supplementary Table 2 enables health care systems to target populations in which lifestyle modifications yield the greatest economic returns.

Less than 10% of the eligible participants (n = 33 of 340; 9.7%) were excluded from the analyses because their Actigraph data was not available. Additionally, data on alcohol consumption and diet were not available in our study. The number of people who scored zero to three on the PLBS was low in our study sample, which may affect the generalizability of the study findings to people with worse lifestyle behaviors. Exploration of the characteristics in individuals with higher annual health care expenditures was also limited by the small subgroup size. Although we adjusted the analyses for participants' report of a recent history of LBP at baseline, and long-term health care cost patterns were captured repetitively for 12 months, reverse causality remains plausible: LBP symptoms may in fact impact lifestyle behaviors (eg, reducing PA or sleep quality), therefore inflating their association with health care costs. Future studies should consider the use of repeated measures of more comprehensive lifestyle factors and include larger samples to further investigate the characteristics and health care use patterns of individuals with lower PLBS and higher health care expenditures associated with LBP. Also, the inclusion of a sample of participants with no history of LBP at baseline who are followed longitudinally until an episode of LBP is observed would be advised, although finding individuals who have never experienced LBP is extremely difficult given the prevalence of the condition.

To conclude, this study demonstrated clear associations between the adoption of positive lifestyle behaviors and a reduction in health care use costs related to LBP. Self-managing and clinical and political interventions aimed at improving the adoption and adherence to positive lifestyle behaviors could potentially yield substantial cost savings for individuals and the health care system while mitigating the burden of LBP.

ACKNOWLEDGMENTS

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney 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, Ms Tian 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. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5:e316–e329.
2. Fatoye F, Gebrye T, Ryan CG, et al. Global and regional estimates of clinical and economic burden of low back pain in high-income countries: a systematic review and meta-analysis. Front Public Health 2023;11:1098100.
3. Back problems. Australian Institute of Health and Welfare. June 17, 2024. Accessed September 22, 2024. https://www.aihw.gov.au/reports/chronic-musculoskeletal-conditions/back-problems
4. Beyera GK, O'Brien J, Campbell S. Determinants of healthcare utilisation for low back pain: a population-based study in Ethiopia. Health Soc Care Community 2020;28:1058–1070.
5. Williams A, van Dongen JM, Kamper SJ, et al. Economic evaluation of a healthy lifestyle intervention for chronic low back pain: a randomized controlled trial. Eur J Pain 2019;23:621–634.
6. AbdulRaheem Y. Unveiling the significance and challenges of integrating prevention levels in healthcare practice. J Prim Care Community Health 2023;14:21501319231186500.
7. National Center for Chronic Disease Prevention and Health Promotion. *The Power of Prevention: Chronic Disease…the Public Health Challenge of the 21st Century*. Centers for Disease Control and Prevention; 2009.
8. Conry MC, Morgan K, Curry P, et al. The clustering of health behaviours in Ireland and their relationship with mental health, self-rated health and quality of life. BMC Public Health 2011;11(1):692.
9. Roberts KE, Beckenkamp PR, Ferreira ML, et al. The impact of aggregate positive lifestyle behaviors on low back pain resilience and care seeking. Spine J 2023;23:1405–1413.
10. Carvalho ESAP, Pinheiro MB, Ferreira ML, et al. Cohort profile: the AUstralian Twin BACK pain and physical activity study (AUTBACK study). BMJ Open 2020;10:e036301.
11. A healthy lifestyle - WHO recommendations. June 20, 2023. Accessed June 11, 2024. World Health Organization. https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle---who-recommendations
12. World Health Organization. WHO guidelines on physical activity and sedentary behaviour. World Health Organization; 2020.
13. Fransson EI, Heikkila K, Nyberg ST, et al. Job strain as a risk factor for leisure-time physical inactivity: an individual-participant meta-analysis of up to 170,000 men and women: the IPD-Work Consortium. Am J Epidemiol 2012;176:1078–1089.
14. Nyberg ST, Singh-Manoux A, Pentti J, et al. Association of healthy lifestyle with years lived without major chronic diseases. JAMA Intern Med 2020;180(5):760–768.
15. Buysse DJ, Reynolds CF III, Monk TH, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989;28(2):193–213.
16. Nieminen LK, Pyysalo LM, Kankaanpaa MJ. Prognostic factors for pain chronicity in low back pain: a systematic review. Pain Rep 2021;6:e919.
17. Heymans MW, van Buuren S, Knol DL, et al. The prognosis of chronic low back pain is determined by changes in pain and disability in the initial period. Spine J 2010;10:847–856.
18. Ferreira ML, Machado G, Latimer J, et al. Factors defining care-seeking in low back pain–a meta-analysis of population based surveys. Eur J Pain 2010;14:e741–e747.
19. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005;44(Pt 2):227–239.
20. Australian Government Department of Health and Aged Care. MBS Online: Medicare Benefits Schedule. Accessed April 11, 2024. https://www9.health.gov.au/mbs/search.cfm
21. Australian Government Department of Health and Aged Care. Pharmaceutical Benefits Scheme (PBS). Accessed May 9, 2024. https://www.pbs.gov.au/pbs/home
22. Belotti F, Deb P, Manning W, et al. twopm: two-part models. Stata J 2015;15(1):3–20.
23. *Stata 18*. Version 18. Stata Statistical Software; 2023. https://www.stata.com/stata18/
24. Walker BF, Muller R, Grant WD. Low back pain in Australian adults. health provider utilization and care seeking. J Manipulative Physiol Ther 2004;27(5):327–335.
25. Redondo-Sendino A, Guallar-Castillon P, Banegas JR, et al. Gender differences in the utilization of health-care services among the older adult population of Spain. BMC Public Health 2006;6:155.
26. Escoffery C. Gender similarities and differences for e-health behaviors among U.S. adults. Telemed J e-Health 2018;24(5):335–343.
27. Celik H, Lagro-Janssen TA, Widdershoven GG, et al. Bringing gender sensitivity into healthcare practice: a systematic review. Patient Educ Couns 2011;84:143–149.
28. Ahern M, Dean CM, Dear BF, et al. The experiences and needs of people seeking primary care for low-back pain in Australia. Pain Rep 2019;4:e756.
29. Altug Z. Lifestyle medicine for chronic lower back pain: an evidence-based approach. Am J Lifestyle Med 2021;15:425–433.
30. Department of Health and Aged Care. National Tobacco Strategy 2023–2030. Department of Health, Disability and Ageing; 2023.
31. Health Ministers Meeting. The National Obesity Strategy 2022–2032. Department of Health, Disability and Ageing; 2022.
32. Breda J, Jewell J, Calvert J, et al. Using Price Policies to Promote Healthier Diets. World Health Organization; 2015.
33. Physical Activity Tax Credit. Department of Finance Government of Newfoundland and Labrador. Accessed December 19, 2024. https://www.gov.nl.ca/fin/tax-programs-incentives/personal/physical-activity-tax-credit/
34. Khow YZ, Lim TLY, Ng JSP, et al. Behavioral impact of national health campaigns on healthy lifestyle practices among young adults in Singapore: a cross-sectional study. BMC Public Health 2021;21:1601.
35. Migueles JH, Cadenas-Sanchez C, Ekelund U, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017;47(9):1821–1845.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 547–556
DOI 10.1002/acr.25648

American College of Rheumatology
Empowering Rheumatology Professionals

Effect of Foot Orthoses on Midfoot Pain and the Volume of Bone Marrow Lesions in the Midfoot: A Randomized Mechanism of Action Study

Jill Halstead,[1,2,3] Anne-Maree Keenan,[3,4] Philip G. Conaghan,[4,5,6] Dennis McGonagle,[1,4] and Anthony C. Redmond[1,4,5,6]

Objective. Foot orthoses are thought to improve pain by potentially modifying internal mechanical forces. To test this, we explored whether foot orthoses can modify patterns of bone marrow lesions (BMLs) in people with midfoot pain.

Methods. Forty-two people were recruited with midfoot pain, and magnetic resonance imaging–confirmed midfoot BMLs. Participants were randomized (2:1 ratio) to receive either pre-formed orthoses (n = 27) or control cushioning insoles (n = 15). Outcomes included foot pain (visual analog scale [VAS]), pain and functional impairment subscales of the Manchester Foot Pain and Disability Index, and BML volume measured at baseline and 12 weeks.

Results. In total 108 bones in the midfoot were identified with BMLs (mean 2.5 bones, SD 1.6). In the orthoses group, pain significantly reduced at 6 weeks (mean VAS = −14.8 mm, confidence interval [CI] −22.3 to −7.3) and 12 weeks (mean VAS = −7.1 mm, CI −15.0 to −0.9) compared to the control group at 6 weeks (mean VAS = −7.4 mm, CI −19.9 to 5.2) and 12 weeks (mean VAS = 2.8 mm, CI −9.1 to 14.7). In the orthoses group, functional impairment and pain impairment were significantly reduced at 6 weeks and to a lesser extent at 12 weeks. In the control group, only the functional impairment reduced significantly at 6 weeks. At 12 weeks, BML volume reduced more in the orthoses group (−1544.4 mm^3, CI −3660.4 to 571.6), compared to the control group (−315.8 mm^3, CI −1528.2 to 896.7).

Conclusion. The foot orthoses group showed a greater reduction in foot pain and a greater reduction in the volume of BMLs compared with the control group.

INTRODUCTION

The prevalence of foot pain has been reported to be between 13% and 36%, and in the midfoot region, the prevalence is 19% over the age of 50 years.[1,2] For people with midfoot pain, walking impairment and disability is similar to midfoot osteoarthritis (OA).[2,3] Foot orthoses are a common treatment for foot pain and have been shown to improve pain, function, and disability in a range of soft tissue and joint disorders.[4] Gait studies show foot orthoses can alter joint motion, change indirect joint forces, and redistribute foot pressure.[5] The mechanism by which foot orthoses improve pain, however, is poorly understood due to the limitations of direct structural measures.

Indirect measures of bone stress and adaptive physiologic changes can be captured using magnetic resonance imaging (MRI), which is sensitive to fluid changes in the bone, defined as bone marrow lesions (BMLs). This nonspecific feature has been shown to reflect physiologic bone changes as well as short-term and long-term adaptive bone stress.[6] Peripheral BMLs have been shown to be responsive to changes in loading and fatigue in the legs and feet,[7,8] associated with increased stress in runners[9] and sports,[10] miliary,[11] and dance participants.[12] These BMLs

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

Supported by a Versus Arthritis studentship (grant 18256). Drs Keenan, Conaghan, McGonagle, and Redmond's work was supported in part by the NIHR Leeds Biomedical Research Centre (grant NIHR203331). The funding source had no role in the study design and collection, analysis, and interpretation of the data; in the writing of the manuscript; or in the decision to submit the manuscript for publication.

[1]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, United Kingdom; [2]Leeds Community Healthcare NHS Trust, Leeds, United Kingdom; [3]School of Healthcare, University of Leeds, Leeds, United Kingdom; [4]NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust, Leeds, United Kingdom; [5]Versus Arthritis Experimental Osteoarthritis Treatment Centre, Leeds, United Kingdom; [6]Centre for Sports, Exercise and Osteoarthritis, Versus Arthritis, Nottingham and Leeds, United Kingdom.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25648).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25648.

Address correspondence via email to Jill Halstead, PhD, at J.Halstead-Rastrick@leeds.ac.uk.

Submitted for publication April 16, 2025; accepted in revised form August 28, 2025.

SIGNIFICANCE & INNOVATIONS

- This is the first clinical study to explore the impact of foot orthoses on magnetic resonance imaging–detected bone marrow lesions in people with midfoot pain.
- Wearing in-shoe arch-contouring firm foot orthoses decreased foot pain over 12 weeks.
- Wearing in-shoe arch-contouring firm foot orthoses also decreased the mean volume of bone marrow lesions across the medial midfoot bones after 12 weeks.

can be a feature of asymptomatic and symptomatic groups undergoing high levels of joint stress, in which the locations are known to be related to repetitive biomechanical activities.[13] In people with pain, BMLs can be reflective of bone stress and underlying OA.[3,14] In people with knee pain, BML size has been shown to correlate weakly with pain, with such pain potentially related to trabecular changes, increased bone vascularity, and perivascular innervation.[14,15,16] The biomechanical relationship between BML and joint and bone pain has been established in animal models of OA[17] and in human knee[14] and hip studies.[18] In the foot, however, this exploration has been limited by a lack of imaging studies and difficulties in tissue sampling. Although our cross-sectional study found the number of people with midfoot pain and bones with BMLs per person was not related,[19] a further study of BMLs in people with midfoot pain and midfoot OA (compared to a control group) showed the size of the BMLs was related to the severity of midfoot pain.[3]

To study the biomechanical adaptation of bone stress in humans, gait studies and imaging studies have explored the effect of orthotic interventions. In the foot, wearing immobilization boots after injury has been shown to resolve BML patterns[20,21] and providing foot orthoses for military training has been shown to prevent bone stress injuries.[22] These foot studies explored orthotic devices and their influence on bone stress in an acute injury model; however the mechanism and role of foot orthoses in a community population with established midfoot pain has not been explored.

To further understand the mechanism of action of foot orthoses, we hypothesized that BMLs at the site of midfoot pain were modifiable by wearing foot orthoses. The aim was to explore whether foot orthoses modified midfoot pain and size of midfoot BMLs over a 12-week period.

PATIENTS AND METHODS

This study had an experimental study design, employing a randomized method to investigate the mechanism of action of foot orthoses and was a substudy of a larger cohort examining the relationship between midfoot pain and BMLs.[19] All eligible participants from the larger study were invited to the orthoses intervention study (International Standard Randomised Controlled Trial Number 77862746).

Participants (aged 18 years or over) were invited from the Leeds Community Musculoskeletal and Rehabilitation service (from 2008 to 2011 in Leeds, United Kingdom) with midfoot pain over three months duration and pain with weight-bearing activity. If participants presented with bilateral pain, the most painful foot was chosen as the study foot, or if the pain was considered equal, a single foot was identified by asking the participants to initiate walking and the first step was determined to identify limb dominance.[19]

Eligible people were invited for an MRI scan on a single foot to meet the inclusion criteria. Two experienced readers (rheumatologist and radiologist) confirmed at least one BML, defined as a hyper-intense signal (visualized on water sensitive T2-weighted fat-saturated or short tau inversion ratio sequences and hypointense signal on T1-weighted sequences) in a region greater than a quarter of one bone in the seven medial midfoot bones (medial cuneiform, intermediate cuneiform, lateral cuneiform, navicular, and proximal first to third metatarsals).

People with midfoot pain were not eligible if they already wore foot orthoses or presented with clinical signs of midfoot OA, defined as observed or palpable osteophytes, reduced range of motion, or identification by radiology report as having features of established midfoot OA using standard clinical foot radiographs (measured in standard anterior–posterior and oblique planes). Other exclusion criteria were foot surgery in the last 12 months and/or internal metal fixation, foot pain typical of undiagnosed or diagnosed inflammatory arthritis, neurologic pain or signs and symptoms of sensory loss and a medical history of diabetes mellitus, peripheral arterial disease, kidney disease, organ transplantation, or being fitted with a pacemaker or any other implant contra-indicated for MRI. In accordance with the Declaration of Helsinki, ethical approval was provided by the Leeds Ethics Committee (reference: 09/H1305/10), and all participants provided informed written consent.

Randomization was conducted on a 2:1 basis favoring the active orthoses intervention over the control (sham) group; this approach was chosen to favor the intervention and explore structural and clinical outcomes. Randomization was provided by senior researcher (AR) using a random number protocol and blind (opaque) envelope allocation, which was revealed to the researcher (JH) sequentially at the baseline appointment. Assessor and participants were not blinded; however, the firm (active) versus cushioning (sham) devices were presented with equal merit to the participants to limit bias.

Participants who were randomized to the intervention group received a pair of Vectorthotic foot orthoses (Healthy Step Ltd): a three-quarter length polypropylene composite shell with rearfoot medial wedges added (either 4° or 6°) and Vectorthotic Extra full length top cover (6 mm to 14 mm variable depth of compressed

polyethylene foam). The control group received the Vectorthotic uniform top cover comprised of 4-mm compressed foam. It was expected that the cushioning sham insole may reduce load in the heel and forefoot to a small degree, whereas the Vectorthotic was chosen to increase midfoot contact force and maximum midfoot force[23,24]; see Figure 1A and B and Supplementary S1 for fitting details.

The primary clinical outcome of this study was foot pain severity in the last 24 hours using a 100-mm visual analog scale[25–27] reported at baseline, 6 weeks, and 12 weeks and MRI-detected BML volume, which was measured at baseline and 12 weeks. The secondary outcome included foot pain–related disability (including functional subscale and the pain subscale), measured using the Modified Manchester Foot Pain and Disability Index (MMFPDI)[28,29] and self-reported global improvement by asking if the foot pain is the same, improved, or worse at the 6-week and 12-week appointment.[30] All outcomes were collected by an unblinded assessor.

The imaging outcomes used an MRI scan at baseline and 12 weeks with a single foot scanned at both timepoints using a Siemens Magnetom Verio (3T) large bore MRI scanner (Siemens Medical Solutions). As described previously,[19] all images were acquired using an eight-channel foot and ankle coil, with the foot placed perpendicular to the ankle and magnetic field (β0). To minimize positional error, all images were measured manually from the scout image to ensure that the foot was perpendicular to the ankle in the sagittal plane and centered over the navicular bone at each visit with the following protocol. T2-weighted fat-saturated sequence parameters were as follows: repetition time (TR): 3,000 to 3,600 milliseconds, echo time (TE): 69 milliseconds, flip angle: 155° to 160°, echo train length: 8, 2 mm slices and 0.4 mm interslice gap, matrix 256 × 256 pixels, field of view (FOV) 150 × 150 mm. Short tau inversion ratio sequence parameters were as follows: TR: 4,500 milliseconds, TE: 31 milliseconds, number of excitations: 2, inversion recovery time: 200 milliseconds, flip angle: 150°, echo train length: 11, 3 mm slices and 0.6 mm interslice gap, matrix 320 × 256 pixels, FOV 150 × 150 mm in all three planes. T1-weighted sequence parameters were spin echo high-resolution, TR: 700 milliseconds, TE: 10 milliseconds, flip angle: 90°, 1.2 mm slices and 1.32 mm interslice gap, matrix 512 × 512 pixels, FOV 512 × 512 mm, in the sagittal plane.

Image analysis. The measurement of bones and BMLs were conducted using anonymized MRIs to ensure the researcher was blinded to group allocation and sequence of scans. Each bone with BMLs was segmented using Analyze software version 10 (AnalyzeDirect). The borders of which were identified using a manual tracing for each cross-sectional area per slice of bone. Once the volume of the bone was determined, the signal of normal bone was identified by manually selecting the lower and upper range of the marrow greyscale signal, which was subtracted from the bone through an automated function leaving the defined BML hyperintense signal (see Supplementary File 2). This software provided two outputs: a cross-sectional slice area of the bone and a cross-sectional area defined as BML signal. With the additional information of slice thickness and the slice gap, the volume of bone and BML per bone was calculated (Volume = $\sum$ Area Bone × [slice thickness + interslice gap]). The individual bone volumes and BML volumes were calculated per scan providing the outcome on a group level and bone level at baseline and follow-up. The percentage BML reduction was also calculated between visits (baseline volume – final volume ÷ baseline volume × 100) on a group level.

The repeatability of this method was investigated in 20 participants; a total of 78 bones measured repeated in a random order over one week later. For the midfoot bones, the absolute difference (root mean squared error) was 1,338.2 mm^3 or 5% of the mean bone volume of the total bones per participant. For BMLs, the absolute difference (root mean squared error) was 427.8 mm^3 or 6.2% of the mean volume of each bone.

Statistical analysis. Differences between the baseline and follow-up scores were reported using the mean and SD, and trends were examined using a two-way repeated measures *t*-test (level of significance was set at $\alpha = 0.05$), and 95% confidence intervals (CIs) were calculated. As this was an exploratory investigation, full inferential analysis was not proposed for this

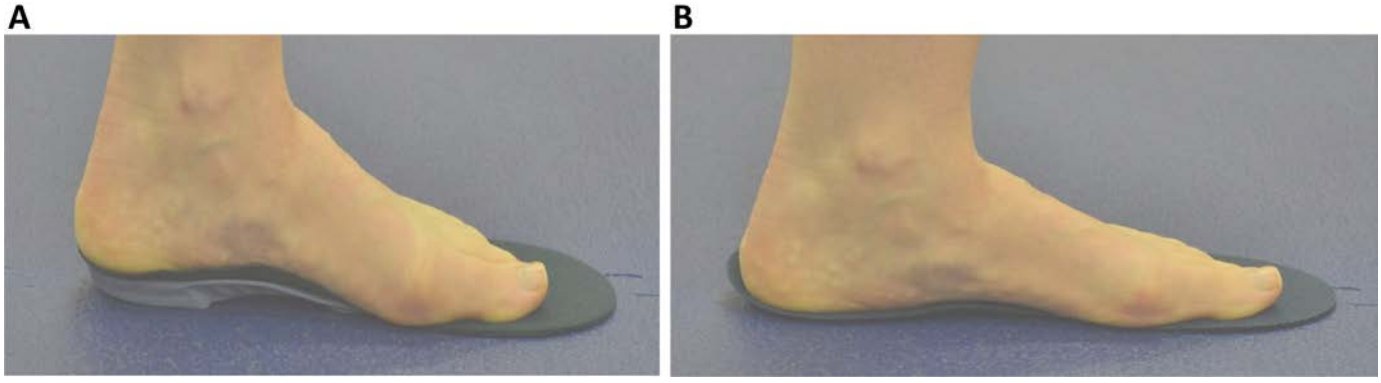

Figure 1. (A) Foot orthoses group—Vectorthotic. (B) Control group—cushioning sham insole.

study. Descriptive analysis of demographic data and outcomes were undertaken using SPSS version 28 (IBM).

RESULTS

Sixty-one people with midfoot pain were invited to have a 3T MRI scan: once the scans were scored, 45 participants with mechanical midfoot pain and confirmed medial midfoot BMLs were invited to the study; 42 were included in the study. Demographics of participants are presented in Table 1: age ranged from 22 to 72 years (mean age 53 [SD 12.2] years), a high proportion were female (71%), and 43% were classified as having obesity and 36% as having overweight (mean body mass index = 30 [SD 5.1]). The most frequently reported comorbidities were OA (n = 17), obesity (n = 13), hypertension (n = 12), and asthma (n = 7), and the most frequently reported sites of other joint pain were the knee (n = 17), hand (n = 7), neck (n = 5), and hip (n = 5). Foot pain was the only site in 33% of the group, whereas 38% reported pain in one other joint, 21% reported pain in two additional joints, and 7% reported pain in three other joints. The mean number of bones with BML per participant were evenly spread, although the mean percentage of BMLs per bone was slightly higher in the control group. The location of bones (n = 108) with BMLs at baseline are shown in Figure 2A. The intermediate cuneiform was the most common bone with BMLs (21%), followed by the second metatarsal bone (19%), and around a third of the group had a pattern of two bones involved.

Randomization. 45 participants were randomized on a 2:1 ratio for treatment with foot orthoses intervention (orthoses group n = 30) or control cushioned insoles (control group n = 15). After randomization, 3 immediately dropped out of the study in the orthoses intervention group, leaving 27 participants. Randomization and the small sample size led to characteristics being unequally distributed: age was higher in the orthoses group and number of painful joints was lower in the control group (see Table 1). An additional five from the orthoses group did not complete the three-month follow-up (see CONSORT statement Supplementary File 3): one participant left the study due to a flare of back pain, three experienced ankle sprains due to icy weather, and one was lost to follow-up.

Intervention. Adherence, number of hours wearing the foot orthoses per day, was evaluated using a daily log. At the six-week review, the mean hours per week was slightly higher in the control group (mean 34.5 [SD 16.6] hours) compared to the orthoses group (mean 32.2 [SD 13.7] hours). At the final review, adherence remained slightly higher in the control group (mean 42.3 [SD 17.0] hours) compared to the orthoses group (mean 39.5 [22.4] hours). Orthoses adherence improved over the course of the three-month study.

Clinical outcomes. In the orthoses group, mean midfoot pain (see Table 2) reduced in the first 6 weeks (mean –14.8 mm, CI –22.3 to –7.3 mm), and smaller mean pain reductions were noted from 6 to 12 weeks (mean –7.1, CI –15.0 to 0.9 mm). In contrast, the control group showed small mean midfoot pain reduction in the first 6 weeks (mean –7.4 mm, CI –19.9 to 5.2 mm) and a small mean increase of foot pain between 6 and 12 weeks (mean 2.8 mm, CI –9.1 to 14.7 mm).

Both groups showed small reductions of functional impairment (MMFPDI) in the first 6 weeks of orthoses therapy (intervention mean –3.5, CI –5.1 to –1.8; control mean –2.1, CI –4.3 to 0.0). There was a greater mean reduction noted in the orthoses intervention group; however, the CIs were similar in both groups. At the 12-week follow-up, both groups showed no change in functional impairment (orthoses mean –0.4, CI –1.7 to 0.9; control mean –0.4, CI –2.4 to 1.7).

Foot pain impairment (MMFPDI) was reduced in the orthoses group in the first 6 weeks (mean –3.5, CI –5.1 to –1.9), although at 12 weeks there was minimal change (mean +0.8, CI –0.3 to 1.9). In contrast, change of foot pain impairment in the control group after 6 weeks was mean –0.7 (CI –1.1 to 2.5) versus mean +0.1 (CI –1.3 to 1.5) for 12 weeks follow-up. These results

Table 1. Demographic profile and baseline characteristics of the total group and randomized orthoses and control treatment groups*

Demographics and baseline characteristics	Total group (n = 42)	Orthoses group (n= 27)	Control group (n = 15)
Age, mean (SD), y	53 (12.2)	55 (10.8)	49 (13.6)
Gender (female), n (%)	30 (71)	20 (77)	10 (63)
BMI, mean (SD)	30.0 (5.1)	29.5 (4.1)	30.6 (6.5)
Total comorbidities, mean (SD)	2.2 (1.6)	2.1 (1.5)	2.4 (1.8)
Total painful joints, mean (SD)	1.4 (1.2)	1.7 (1.2)	0.8 (0.9)
Pain (0–100 mm), mean (SD)	34.6 (17.3)	36.5 (16.1)	32.2 (19.4)
Pain related disability, mean (SD)	30.5 (6.1)	30.6 (5.8)	30.5 (6.8)
Pain impairment, mean (SD)	13.6 (3.0)	14.1 (2.7)	12.9 (3.3)
Functional impairment, mean (SD)	16.9 (4.6)	16.5 (4.7)	17.6 (4.6)
No. of bones with BMLs, mean (SD)	2.6 (1.6)	2.6 (1.5)	2.6 (1.9)
BML volume % of bone, mean (SD)	35.1 (19.5)	31.3 (19.5)	35.1 (19.8)
Study foot	26 right	15 right	11 right

* BMI, body mass index, BML, bone marrow lesion.

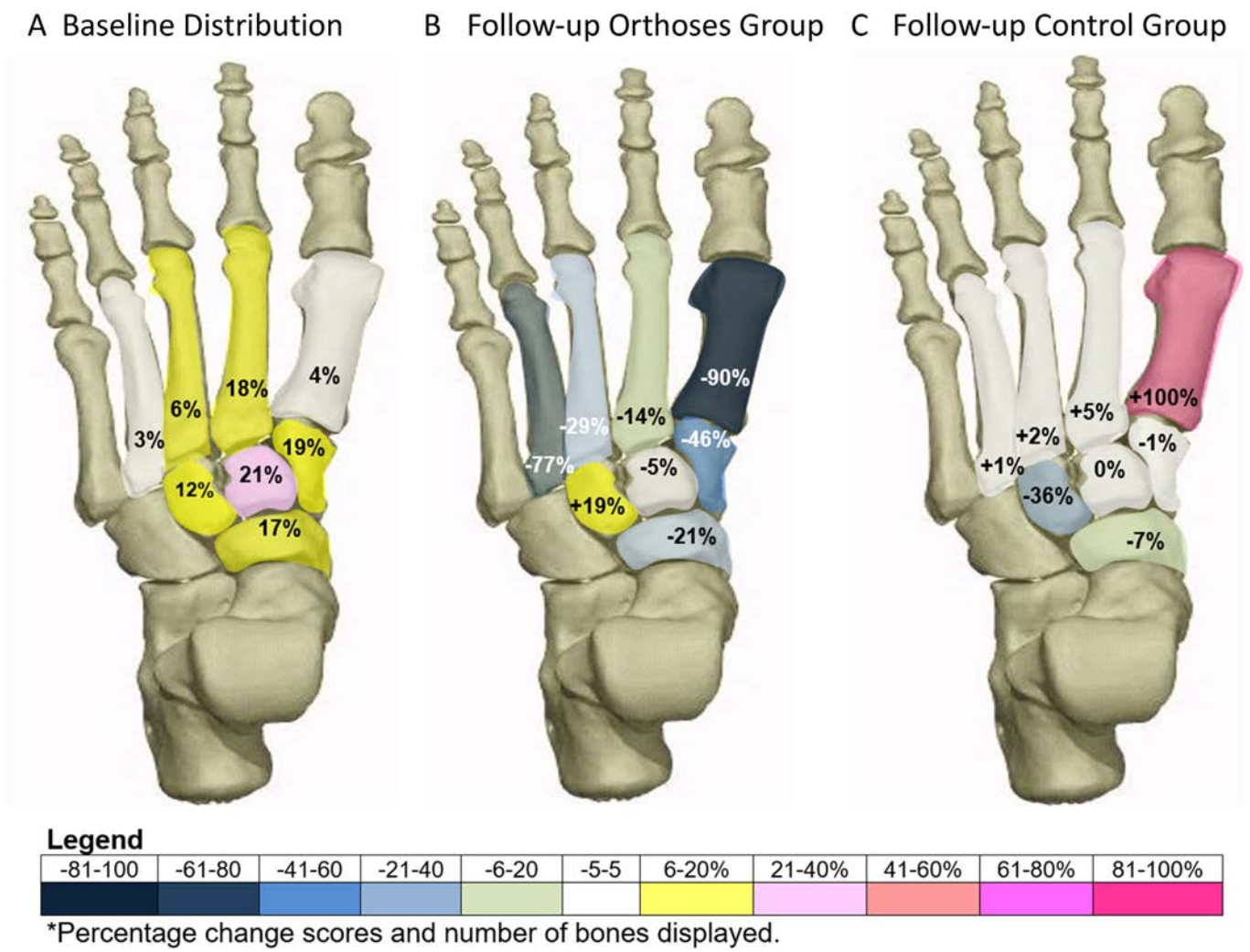

Figure 2. (A) shows the proportion of bones with reported BMLs at baseline as a total group. (B) and (C) show the 12-week BML percentage changes per bone for the (B) orthoses group and (C) control group. BML, bone marrow lesion. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25648/abstract.

suggest midfoot pain and foot pain impairment reduced in the first 6 weeks in the orthoses group and not in the control group. Functional impairment reduced in both groups in the first 6 weeks of intervention and then plateaued. In the self-reported evaluation of the pain (see Table 2), 86% of participants in the orthoses group reported improvement, compared to 40% of the control group, after 12 weeks.

Structural outcomes. After 12 weeks, the baseline and follow-up BML patterns were explored in each group in terms of total mean measurements for bone volume, BML volume, and BML percentage per bone (see Table 3). Mean bone volumes were similar in both groups, with some increase in bone volume measurement in the control group at follow-up due to two more bones being detected. Although both groups showed a reduction in BML volume after 12 weeks, a larger reduction was noted in the orthoses group by a mean of –1544.4 mm^3 (CI –3660.4 to 571.6 mm^3) with an overall percentage decrease of 25.8% and percentage reduction per bone of –6.6% (CI –14% to 0.8%). In the control group, there was a smaller mean reduction of –315.8 mm^3 in BML volume (CI –1528.2 to 896.7 mm^3), with an overall percentage decrease of 4.4% and a percentage reduction per bone of –0.5% (CI –6.6% to 5.8%).

New BML formation was noted in both groups at 12 weeks. In the control group, four participants reported new BML formation in varying degrees of size in the navicular, medial cuneiform, intermediate cuneiform, and first metatarsal bones (5%, 6%, 11%, and 100% of bone, respectively). In the orthoses group, one participant had new BMLs in three bones: lateral (58%), medial (17%), and intermediate (41%) cuneiforms. The other three BMLs were newly formed in three intermediate cuneiforms (64%, 6%, and 6% of bone). None of the BML resolutions occurred in the same participants with new BML formation.

The BML results were further explored on bone level for the relative change in number of bones and percentage BML volume per group (see Table 4 and Figure 2B and 2C). Figure 2 presents the

Table 2. Mean foot pain (VAS) and foot pain impairment and functional impairment (MMFPDI) and self-reported improvement at baseline and follow-up in each group*

Outcomes	Orthoses group						Control group					
	Week 0 (n = 27), mean (SD)	Week 6 (n = 25), mean (SD)	Week 12 (n = 22), mean (SD)	Week 0 to 6, difference (CI)	Week 6 to 12, difference (CI)	Week 0 to 12,[a] difference (CI)	Week 0 (n = 15), mean (SD)	Week 6 (n = 15), mean (SD)	Week 12 (n = 15), mean (SD)	Week 0 to 6, difference (CI)	Week 6 to 12, difference (CI)	Week 0 to 12, difference (CI)
Foot pain	35.7 (16.2)	20.5 (17.5)	14.6 (17.9)	−14.8 (−22.3 to −7.3)	−6.0 (−15.0 to 0.9)	−18.5 (−28.5 to −8.5)	32.7 (20.0)	24.3 (22.3)	24.9 (19.5)	−7.4 (−19.9 to 5.2)	2.8 (−9.1 to 14.7)	−7.8 (−21.2 to 5.6)
Foot pain impairment	14.0 (2.7)	10.6 (3.3)	10.8 (3.0)	−3.5 (−5.1 to −1.9)	0.8 (−0.3 to 1.9)	−2.7 (−4.3 to −1)	13.0 (3.4)	11.9 (2.9)	12.1 (3.2)	−0.7 (−1.1 to 2.5)	0.1 (−1.3 to 1.5)	−0.9 (−2.4 to 0.7)
Foot function impairment	16.5 (4.7)	13.6 (4.8)	13.0 (4.2)	−3.5 (−5.1 to −1.8)	−0.4 (−1.7 to 0.9)	−3.5 (−5.2 to −1.9)	17.5 (4.8)	15.2 (4.5)	14.9 (1.7)	−2.1 (−4.3 to 0.0)	−0.4 (−2.4 to 1.7)	−2.7 (−4.0 to −1.3)
Self-reported improvement, %												
Improved	–	76	86	–	–	–	–	53	40	–	–	–
Same	–	16	9	–	–	–	–	40	47	–	–	–
Worse	–	8	5	–	–	–	–	7	13	–	–	–

* CI, 95% confidence interval; MMFPDI, Modified Manchester Foot Pain and Disability Index; VAS, visual analog scale.
[a] Excluding 5 participants who did not complete the study in the orthoses group.

distribution of bones with BMLs across the midfoot at baseline, which demonstrates that the intermediate cuneiform, medial cuneiform, navicular, and second metatarsal were most affected. It also shows the percentage change of BML volume per bone per intervention and control group at 12 weeks. In the orthoses group, a total of seven bones showed complete resolution of BML, and the remaining seven of the eight BML bone sites showed some reduction, whereas in the control group, two bones with BML resolved, and five of the eight BML bone sites showed some BML reduction. In the case of the medial cuneiform and the intermediate cuneiform bones, there were increases and decreases in BML volume. In the orthoses intervention group, the mean percentage bone reductions were in the medial cuneiform, navicular, and metatarsal bones. In contrast, these bones showed very little mean change in the control group.

DISCUSSION

This study aimed to investigate the mechanism of action of foot orthoses in people with midfoot pain and walking difficulties, using clinical outcomes and MRI scans to measure the effect on BMLs as a surrogate measure of bone stress. In this novel study, midfoot pain and MRI-detected bone signal in the midfoot were assessed before and after 12 weeks of wearing either firm foot orthoses or cushioning sham insole. The results of this exploratory study showed that average foot pain reduced in both groups, but there was a greater reduction in the orthoses intervention group, and more people reported improvement after 12 weeks. In terms of MRI-detected bone signal, after the intervention, total mean volume of BML and total mean percentage of BML per bone was lower in the orthoses intervention group than the control group at 12 weeks. In the intervention group, the reduction of BML size and foot pain occurred over a shorter period than previously reported hospital studies, which showed foot BMLs took six to nine months to reduce[31] and up to one year for resolution.[32] Pain reduction in conjunction with BML reduction suggests further studies are needed to explore the role of bone stress in midfoot pain.

In the orthoses group, the BML volume bone reductions were mostly found in the medial cuneiform, navicular, and metatarsal bones. In contrast, these same bones showed very little mean change in the control group, suggesting the orthoses may influence the distribution of the mechanical stress within the medial midfoot. On an individual bone level in the orthoses group, a greater number of bones showed either resolution of BMLs or percentage BML reductions, compared to the control group. In the case of the medial cuneiform, and particularly the intermediate cuneiform, the results suggested these bones were sensitive to changes in bone stress.

Bone marrow signal detected on MRI scans is a measure of abnormal physiology[33] and, in the foot BMLs, are associated with high repetitive stress, such as excessive training, in running, sports, and the miliary.[34] In a clinical study of hospital MRI scans in people with foot and ankle pain, BMLs were primarily

Table 3. The mean difference between 0 and 12 weeks follow-up BML outcomes per group*

Bone and BML measurements	Orthoses group Week 0[a]	Orthoses group Week 12	Difference (CI) n = 22	Control group Week 0	Control group Week 12	Difference (CI) n = 15
Bone volume, mean (SD) mm^3	18,380.5 (9,828.4)	18,300.5 (9,453.4)	−79.9 (−556.7 to 396.8) $P = 0.730$	19,750.5 (15,763.3)	20,234.7 (16,269.2)	484.2 (−51.9 to 1,020.3) $P = 0.73$
BML volume, mean (SD) mm^3	5,977.3 (5,023.0)	4,432.9 (3,239.5)	−1,544.4 (−3,660.4 to 571.6) $P = 0.144$	7,183.8 (9,125.4)	6,868.1 (8,533.3)	−315.8 (−1,528.2 to 896.7) $P = 0.587$
BML per bone, n (%)	33.2 (20.8)	26.6 (19.5)	−6.6 (−14.0 to 0.8) $P = 0.078$	35.1 (19.8)	34.6 (24.3)	−0.5 (−6.6 to 5.8) $P = 0.892$

* *P* is the probability of 2-tailed *t*-test, significance set at 5%. BML, bone marrow lesion, CI, 95% confidence interval.
[a] Excluding 5 participants who did not complete the study in the orthoses group.

associated with trauma (44%), OA (35%), and primary BML (6%), followed by (<5%) infection, ischemia, and neoplasm.[35] In this community study, people were recruited with an average duration of 10 months foot pain and a pattern of weight-bearing foot pain that was not related to trauma, and people with clinical and radiographic signs of OA were excluded. The baseline MRI characteristics of people with midfoot pain, as we have previously published, show that the number of bones with BMLs was not related to pain and that these features were reported alongside MRI-detected features of OA (mainly joint space narrowing and osteophytes, which are poorly visualized in the small joints on x-ray) and that this trend increased with age,[19] in keeping with previous studies.[36] In a second study we have published, when comparing a control group (with no foot pain) to a midfoot pain group and radiographic confirmed midfoot OA group, the size of the BMLs was associated with midfoot pain.[3] This suggests the possible mechanism of foot orthoses to modify foot BMLs may be applied to stress-related and OA-related BMLs.

Using devices to modify or reduce bone stress has been explored to reduce injury and improve pain. Immobilization therapy is perhaps the most well-known treatment for acute injuries of the foot and ankle, and walker boots or casts have been shown to reduce and resolve BMLs over one to four months.[20,21] Outside of this treatment, less in known about the biomechanical properties of in-shoe devices and their influence on bone stress. An important experimental study of functional foot orthoses reported that an immediate reduction of bone stress (using an invasive bone-mounted strain-gauges) was evidenced in the tibial and second metatarsal bones upon initiation of orthotic use.[37,38] This is supported by a finite element imaging model of midfoot OA, showing navicular and cuneiform cartilage and bone stress was modified by changes in orthoses stiffness and arch height.[39] The response of bone to external device has also been shown by wearing an in-shoe pad. After running for two weeks with a pad placed on the lateral midfoot (aimed to induce foot pronation), baseline and follow-up MRI scans showed BMLs occurred in lateral foot bones in 10 of 11 cases.[40] More clinical studies have followed to test whether orthotic devices fitted to support the foot during running and intense training can reduce or prevent bone

Table 4. Frequency and change of BMLs per bone site for the total group and each treatment arm: orthoses and control groups at 0 and 12 weeks follow-up*

Groups	Total group	Orthoses group				Control group			
			Weeks 12 BML change (n = 22)				Week 12 BML change (n = 15)		
Time point (participants)	Baseline (N = 42)	Week 0 (n = 22)[a]		Bone change, n	Mean BML vol change,[b] %	Week 0 (n = 15)		Bone change, n	Mean BML vol change, %
Total bones	108	54	53			42	44		
Bone sites, n (%)									
1st metatarsal	4 (4)	2 (4)	1 (2)	−1	−90	0 (0)	1 (2)	+1	+100
2nd metatarsal	20 (19)	10 (18)	9 (17)	−1	−15	8 (19)	8 (18)	0	+5
3rd metatarsal	6 (6)	3 (6)	3 (6)	0	−28	3 (7)	3 (7)	0	−2
4th metatarsal	3 (3)	1 (2)	1 (2)	0	−77	2 (5)	2 (5)	0	+1
Medial cuneiform	20 (19)	11 (20)	11 (20)	+1/−1	−46	6 (14)	6 (14)	+1/−1	−7
Intermediate cuneiform	23 (21)	11 (20)	11 (20)	+4/−4	−4	9 (21)	9 (20)	+1/−1	0
Lateral cuneiform	14 (12)	6 (11)	7 (13)	+1	+24	7 (17)	7 (16)	0	−36
Navicular	18 (18)	10 (18)	10 (19)	0	−25	7 (17)	8 (18)	+1	−1

* BML, bone marrow lesion; vol, volume.
[a] Excluding 5 participants who did not complete the study in the orthose.
[b] Mean BML change is the percentage reduction per bone.

stress. In a military randomized controlled sham trial, in-shoe foot orthoses reduced foot and lower leg injuries more than the sham group[23] and, specifically, the incidence of medial tibial stress was lower.[41] These studies support the gait studies, which suggest that foot orthoses can alter foot biomechanics,[42] and it shows the potential of orthoses to support bone adaptation and prevent injury. In contrast, running shoes designed to simulate barefoot running and reduce support of the foot show increase in BMLs in the metatarsals.[43] These studies show the potential for BMLs to be induced and modified in a relatively short period (from two to seven weeks).

Taking into account the previous research, this novel exploratory study suggests that wearing orthoses for 12 weeks can reduce midfoot pain and BML size compared to the control sham group. The orthoses intervention also reduced short-term pain and impairment more than the control insole. The reductions in pain were similar to those reported in plantar fasciitis and posterior tibial tendon dysfunction randomized trials.[3,4,44–46] These results in the foot also align with a knee brace study, which also showed knee pain and patellofemoral BML size were reduced in people with OA.[47] This is important as longitudinal studies of the knee and foot suggest there is a relationship between the size of the BML, higher pain levels,[48] and longer duration of pain,[32] suggesting reducing pain related midfoot BMLs may have long-term benefits and further studies are needed.

This exploratory study had limitations, especially the size of the study, that limited assumptions and generalizability, as a sample size calculation was not performed. Due to the small sample and number of midfoot bones, an unequal randomization method (2:1 ratio) was chosen, favoring greater numbers in the intervention group, as one of the main concerns was BML migration in the intervention group, which has been show in a one-year follow-up study.[49] This therefore limited direct comparison between the intervention and control groups. The outcome measure to assess pain and impairment MMFPDI may have also limited the study, as both groups showed similar reductions over 12 weeks despite differences in overall patient-reported improvement. This is in agreement with a footwear intervention study, in which the responsiveness of the MMFPDI was limited in an older population with foot pain.[50]

This study results may also be affected by response bias from both the researcher and participant, which may lead to greater positive results in the intervention group. To limit bias, the researcher presented each intervention device equally to the participants; however, patient beliefs were not fully assessed, and this would be recommended for future studies. The MRI results were analyzed anonymously without knowledge of baseline, follow-up, and intervention group status to limit bias. Finally, this study set out to exclude people with radiographic OA, which may be less responsive to modification of BMLs over a short period. The baseline semiquantitative scoring showed MRI-detected BMLs were detected alongside joint changes typical of OA,[19] which may have affected the results.

In conclusion, foot orthoses are widely prescribed for foot pain, however the mechanism of action is not fully understood. To date, the physiologic effect of foot orthoses to mediate internal foot forces requires further research in clinical community populations, particularly in relation to bone stress. This exploratory study addresses some of this gap and provides data that shows that patient improvement, reduction in midfoot pain, and reduction in BML size were greater in the foot orthoses group than the control sham group.

ACKNOWLEDGMENTS

The authors would like to thank all the participants for providing their time to support this study. The authors would also like to acknowledge members of the Biomedical Research Centre imaging team, including Dr Richard Hodgson, Rob Evans, and Dr Carole Burnett, for the acquisition of the MRI scans for the project.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Halstead 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. Gates LS, Arden NK, Hannan MT, et al. Prevalence of foot pain across an international consortium of population-based cohorts. Arthritis Care Res (Hoboken) 2019;71(5):661–670.
2. Thomas MJ, Peat G, Rathod T, et al. The epidemiology of symptomatic midfoot osteoarthritis in community-dwelling older adults: cross-sectional findings from the Clinical Assessment Study of the Foot. Arthritis Res Ther 2015;17(1):178.
3. Arnold JB, Halstead J, Martín-Hervás C, et al. Bone marrow lesions and magnetic resonance imaging-detected structural abnormalities in patients with midfoot pain and osteoarthritis: a cross-sectional study. Arthritis Care Res (Hoboken) 2023;75(5):1113–1122.
4. Banwell HA, Mackintosh S, Thewlis D. Foot orthoses for adults with flexible pes planus: a systematic review. J Foot Ankle Res 2014; 7(1):23.
5. Mills K, Blanch P, Chapman AR, et al. Foot orthoses and gait: a systematic review and meta-analysis of literature pertaining to potential mechanisms. Br J Sports Med 2010;44(14):1035–1046.
6. Koushesh S, Shahtaheri SM, McWilliams DF, et al. The osteoarthritis bone score (OABS): a new histological scoring system for the characterisation of bone marrow lesions in osteoarthritis. Osteoarthritis Cartilage 2022;30(5):746–755.
7. Mandell JC, Khurana B, Smith SE. Stress fractures of the foot and ankle, part 1: biomechanics of bone and principles of imaging and treatment. Skeletal Radiol 2017;46(8):1021–1029.
8. Cowley E, Marsden J. The effects of prolonged running on foot posture: a repeated measures study of half marathon runners using the foot posture index and navicular height. J Foot Ankle Res 2013; 6(1):20.

9. Freund W, Weber F, Billich C, et al. The foot in multistage ultramarathon runners: experience in a cohort study of 22 participants of the Trans Europe Footrace Project with mobile MRI. BMJ Open 2012;2(3):e001118.
10. Tenforde AS, Outerleys J, Bouxsein ML, et al. Metatarsal bone marrow edema on magnetic resonance imaging and its correlation to bone stress injuries in male collegiate basketball players. Orthop J Sports Med 2022;10(1):23259671211063505.
11. Greeves JP, Beck B, Nindl BC, et al. Current risks factors and emerging biomarkers for bone stress injuries in military personnel. J Sci Med Sport 2023;26(suppl 1):S14–S21.
12. Miskovsky S, Khambete P, Faraji N, et al. Prevalence of asymptomatic talar bone marrow edema in professional ballet dancers: preliminary data from a 2-year prospective study. Orthop J Sports Med 2023; 11(5):23259671231159910.
13. Katakura M, Clark R, Lee JC, et al. Foot and ankle MRI findings in asymptomatic professional ballet dancers. Orthop J Sports Med 2024;12(8):23259671241263593.
14. Barr AJ, Campbell TM, Hopkinson D, et al. A systematic review of the relationship between subchondral bone features, pain and structural pathology in peripheral joint osteoarthritis. Arthritis Res Ther 2015; 17(1):228.
15. Zanetti M, Bruder E, Romero J, et al. Bone marrow edema pattern in osteoarthritic knees: correlation between MR imaging and histologic findings. Radiology 2000;215(3):835–840.
16. Aso K, Sugimura N, Wada H, et al. Increased nerve growth factor expression and osteoclast density are associated with subchondral bone marrow lesions in osteoarthritic knees. Osteoarthr Cartil Open 2024;6(3):100504.
17. Hansen RT, Chenu C, Sofat N, et al. Bone marrow lesions: plugging the holes in our knowledge using animal models. Nat Rev Rheumatol 2023;19(7):429–445.
18. Fang H, Zhang X, Wang J, et al. The relationship between MRI-detected hip abnormalities and hip pain in hip osteoarthritis: a systematic review. Rheumatol Int 2024;44(10):1887–1896.
19. Halstead J, Martín-Hervás C, Hensor EMA, et al. Association between clinical and MRI-detected imaging findings for people with midfoot pain, a cross-sectional study. J Foot Ankle Res 2025;18(1):e70019.
20. Elias I, Zoga AC, Schweitzer ME, et al. A specific bone marrow edema around the foot and ankle following trauma and immobilization therapy: pattern description and potential clinical relevance. Foot Ankle Int 2007;28(4):463–471.
21. Aigner N, Radda C, Meizer R, et al. Bone marrow edema in the foot - MRI findings after conservative therapy. Foot Ankle Surg 2005;11(2): 87–91.
22. Bonanno DR, Murley GS, Munteanu SE, et al. Effectiveness of foot orthoses for the prevention of lower limb overuse injuries in naval recruits: a randomised controlled trial. Br J Sports Med 2018;52(5): 298–302.
23. McCormick CJ, Bonanno DR, Landorf KB. The effect of customised and sham foot orthoses on plantar pressures. J Foot Ankle Res 2013;6(1):19.
24. Chapman GJ, Halstead J, Redmond AC. Comparability of off the shelf foot orthoses in the redistribution of forces in midfoot osteoarthritis patients. Gait Posture 2016;49:235–240.
25. de Boer AG, van Lanschot JJ, Stalmeier PF, et al. Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life? Qual Life Res 2004;13(2):311–320.
26. Averbuch M, Katzper M. Assessment of visual analog versus categorical scale for measurement of osteoarthritis pain. J Clin Pharmacol 2004;44(4):368–372.
27. Pincus T, Bergman M, Sokka T, et al. Visual analog scales in formats other than a 10 centimeter horizontal line to assess pain and other clinical data. J Rheumatol 2008;35(8):1550–1558.
28. Garrow AP, Papageorgiou AC, Silman AJ, et al. Development and validation of a questionnaire to assess disabling foot pain. Pain 2000; 85(1–2):107–113.
29. Cook C, Cleland J, Pietrobon R, et al. Calibration of an item pool for assessing the disability associated with foot pain: an application of item response theory to the Manchester Foot Pain and Disability Index. Physiotherapy 2007;93(2):89–95.
30. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther 2009;17(3):163–170.
31. Tonbul M, Guzelant AY, Gonen A, et al. Relationship between the size of bone marrow edema of the talus and ankle pain. J Am Podiatr Med Assoc 2011;101(5):430–436.
32. Zanetti M, Steiner CL, Seifert B, et al. Clinical outcome of edema-like bone marrow abnormalities of the foot. Radiology 2002;222(1): 184–188.
33. Thiryayi WA, Thiryayi SA, Freemont AJ. Histopathological perspective on bone marrow oedema, reactive bone change and haemorrhage. Eur J Radiol 2008;67(1):62–67.
34. Tarantino U, Greggi C, Cariati I, et al. Reviewing bone marrow edema in athletes: a difficult diagnostic and clinical approach. Medicina (Kaunas) 2021;57(11):1143.
35. González-Martín D, Herrera-Pérez M, Martín-Vélez P, et al. Prevalence of bone marrow edema in a study population with foot and/or ankle pain. Foot (Edinb) 2019;40:76–80.
36. Zubler V, Mengiardi B, Pfirrmann CWA, et al. Bone marrow changes on STIR MR images of asymptomatic feet and ankles. Eur Radiol 2007;17(12):3066–3072.
37. Meardon SA, Edwards B, Ward E, et al. Effects of custom and semi-custom foot orthotics on second metatarsal bone strain during dynamic gait simulation. Foot Ankle Int 2009;30(10):998–1004.
38. Ekenman I, Milgrom C, Finestone A, et al. The role of biomechanical shoe orthoses in tibial stress fracture prevention. Am J Sports Med 2002;30(6):866–870.
39. Zhang H, Lin Lv M, Yang J, et al. Computational modelling of foot orthosis for midfoot arthritis: a Taguchi approach for design optimization. Acta Bioeng Biomech 2020;22(4):75–83.
40. Schweitzer ME, White LM. Does altered biomechanics cause marrow edema? Radiology 1996;198(3):851–853.
41. Franklyn-Miller A, Wilson C, Bilzon J, et al. Foot orthoses in the prevention of injury in initial military training: a randomized controlled trial. Am J Sports Med 2011;39(1):30–37.
42. Jafarnezhadgero A, Esmaeili A, Hamed Mousavi S, et al. Effects of foot orthoses application during walking on lower limb joint angles and moments in adults with flat feet: a systematic review with meta-analysis. J Biomech 2024;176:112345.
43. Ridge ST, Johnson AW, Mitchell UH, et al. Foot bone marrow edema after a 10-wk transition to minimalist running shoes. Med Sci Sports Exerc 2013;45(7):1363–1368.
44. Kulig K, Reischl SF, Pomrantz AB, et al. Nonsurgical management of posterior tibial tendon dysfunction with orthoses and resistive exercise: a randomized controlled trial. Phys Ther 2009;89(1):26–37.
45. Landorf KB, Keenan AM, Herbert RD. Effectiveness of foot orthoses to treat plantar fasciitis: a randomized trial. Arch Intern Med 2006; 166(12):1305–1310.
46. Baldassin V, Gomes CR, Beraldo PS. Effectiveness of prefabricated and customized foot orthoses made from low-cost foam for noncomplicated plantar fasciitis: a randomized controlled trial. Arch Phys Med Rehabil 2009;90(4):701–706.

47. Callaghan MJ, Parkes MJ, Hutchinson CE, et al. A randomised trial of a brace for patellofemoral osteoarthritis targeting knee pain and bone marrow lesions. Ann Rheum Dis 2015;74(6): 1164–1170.
48. Zhang Y, Nevitt M, Niu J, et al. Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. Arthritis Rheum 2011;63(3):691–699.
49. Fernandez-Canton G, Casado O, Capelastegui A, et al. Bone marrow edema syndrome of the foot: one year follow-up with MR imaging. Skeletal Radiol 2003;32(5):273–278.
50. Menz HB, Auhl M, Ristevski S, et al. Comparison of the responsiveness of the foot health status questionnaire and the Manchester foot pain and disability index in older people. Health Qual Life Outcomes 2014;12(1):158.

Arthritis Care & Research
Vol. 78, No. 4, April 2026, pp 557–560

American College of Rheumatology
Empowering Rheumatology Professionals

LETTER

DOI 10.1002/acr.25666

Gut–lung axis and systemic sclerosis–associated interstitial lung disease: toward validated multimodal predictive models: comment on the article by Andréasson et al

To the Editor:

We would like to draw your attention to the article "International investigation of the gut–lung axis in systemic sclerosis–interstitial lung disease" by Andréasson et al. By integrating shotgun metagenomics with quantitative high-resolution computed tomography (HRCT) analysis, the authors report a cohort of 285 patients with SSc (62.5% of whom have interstitial lung disease [ILD]) and identify a multivariate intestinal species signature that discriminates ILD and associates with radiologic severity. These findings substantiate a species-level link between the gut microbiome and ILD burden in SSc and lay a concrete foundation for translational applications.[1]

Recognizing the exploratory nature of these analyses, we encourage two methodologic refinements to accelerate clinical translation: (1) the routine application of multiple-testing control (false discovery rate [FDR]) to taxonomic and functional features and (2) site-robust validation (eg, leave-one-site-out) to probe generalizability across centers. In parallel, prospective substudies with synchronous biospecimen collection (stool and serum) and HRCT performed in tight temporal proximity would minimize timing confounding and strengthen microbe-phenotype inferences.

On the translational front, we see immediate opportunities for multimodal risk models that combine the intestinal signature with circulating and cardiopulmonary markers. In SSc-ILD, longitudinal change in Krebs von den Lungen 6 (KL-6) has emerged as a treatment-response biomarker and improves model discrimination, supporting its integration with microbiome-based predictors.[2] A cardiopulmonary dimension is likewise pertinent: in SSc with pulmonary hypertension, a reduced tricuspid annular plane systolic excursion (TAPSE) and systolic pulmonary artery pressure (sPAP) ratio independently marks higher risk of death and cardiovascular events, suggesting that right-heart coupling and ILD should be comodeled in comprehensive prognostic tools.[3] Immune profiling could further refine such models: LOXHD1 and RHOB expression in monocytes has been associated with progressive SSc-ILD, offering mechanistic complementarity to KL-6 and C-reactive protein and to the microbial signature.[4]

The broader relevance of a "multi-axis" approach is echoed outside SSc. In primary Sjögren disease with ILD, patients display greater cardiopulmonary vulnerability (eg, higher sPAP and lower TAPSE:sPAP ratio) than counterparts without ILD, arguing for structured echocardiographic monitoring and integrative phenotyping.[5] To reduce radiation burden and enable dense monitoring, lung ultrasound (LUS) is an attractive adjunct: recent work validating LUS-ILD-24 criteria in SSc and inflammatory myopathies supports expert-interpreted LUS for screening and treatment-response assessment within multimodal pathways.[6]

In summary, Andréasson et al provide compelling evidence that a species-level intestinal signature is associated with both presence and severity of ILD in SSc. We advocate for the following: FDR-controlled, site-robust validation; synchronized HRCT-stool-serum sampling; and multimodal modeling that integrates the microbiome (taxonomic and functional) with KL-6, monocyte transcriptomics, and cardiopulmonary phenotyping (including LUS). Such a framework could enable pragmatic, adaptive trials, curb repeated computed tomography exposure, and accelerate truly personalized strategies for patients with SSc-ILD.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25666.

Gianluca Pagnoni, MD
dr.gianlucapagnoni@gmail.com
Cardiology Unit of Emergency Department,
Guglielmo da Saliceto Hospital
Piacenza, Italy
Aurora Vicenzi, DS
National Institute for Cardiovascular Research
Bologna, Italy
Francesca Coppi, MD
Department of Medical and Surgical Sciences for Children and Adults, University of Modena and Reggio Emilia
Modena, Italy

1. Andréasson K, Young A, Joshi S, et al. International investigation of the gut–lung axis in systemic sclerosis–interstitial lung disease. Arthritis Care Res (Hoboken) 2026:78(3):352–361.
2. Volkmann ER, Wilhalme H, Tashkin DP, et al. Treatment response biomarkers for systemic sclerosis–associated interstitial lung disease. Arthritis Care Res (Hoboken) 2025;77(6):753–759.

3. de Pinto M, Coppi F, Spinella A, et al. The predictive role of the TAPSE/sPAP ratio for cardiovascular events and mortality in systemic sclerosis with pulmonary hypertension. Front Cardiovasc Med 2024;11:1430903.
4. Padilla CM, Lafyatis R, Gibson KF, et al. Monocyte LOXHD1 and RHOB expression predictive of progressive systemic sclerosis–associated interstitial lung disease. Arthritis Care Res (Hoboken) 2026;78(4): 456–468.
5. Coppi F, Pagnoni G, Campani C, et al. Sjögren disease and pulmonary hypertension: exploring the intricate link with interstitial lung disease. Int J Cardiol 2025;430:133185.
6. Fairchild RM, Mar DA, Deluna MD, et al. Validation of lung ultrasound interpretation criteria for interstitial lung disease in systemic sclerosis and inflammatory myopathy. Arthritis Care Res (Hoboken) 2025; 77(11):138–1392.

DOI 10.1002/acr.25667

Reply

To the Editor:

We are grateful for the interest from Pagnoni et al in our study.[1] Indeed, this exploratory investigation is, to our knowledge, the first multicenter study linking bacterial species to systemic sclerosis–related interstitial lung disease (SSc-ILD). In addition, it is the first study in the field of systemic sclerosis to adjust for multiple known confounders in analyses of the intestinal microbiota.

We acknowledge the methodologic issues raised by Pagnoni et al, including suggestions on how to probe generalizability across study centers, how to synchronize biospecimen collection, and statistical methodology. Great technical advancements have been made during the last decade in this field, and we appreciate these comments for our future work. We also note other novel approaches on how to integrate whole genome shotgun sequencing with metabolomics and clinical data.[2]

We recognize the concern from Pagnoni et al regarding multiple comparisons in our species-level association analyses. Because this is the first study of its kind, we employed a discovery-based approach emphasizing effect size estimation and hypothesis generation as compared to hypothesis testing. In the context of microbiome research, wherein large sample sizes are required to detect effects, especially under conditions of multiple hypothesis testing, the results suggest potential targets for future hypothesis-driven analyses. In addition, Cohen's *d* was provided to facilitate sample size estimations needed to guide future investigations in this emerging area of research.

To deal with generalizability of effects, we specified a general linear model with a random effect for sites. This model accounts for the variability in outcomes due to differences in site. The estimate for groups is the weighted average of the within-site and between site components of the group differences. Although this model provides the statistical basis for generalization of results across sites, we agree that the leave-one-out sensitivity analyses is an important tool for assessing the influence of each site on the results and the robustness of statistical models. We acknowledge that there may be important site-specific differences or associations with microbiomes because of diversity of diets and environmental exposures between international sites that the random effects model does not capture. Ultimately, replication in well-powered independent datasets is required to confirm the validity and generalizability of the findings.

In agreement with Pagnoni et al, we welcome future integrative "multiaxis" approaches, possibly incorporating disease phenotyping based on circulating biomarkers, immune profiling of leukocytes, and echocardiological assessments. In our study, we used quantitative image analysis of high-resolution computed tomography,[3] a valid approach for approximating the burden of pulmonary fibrosis that is more sensitive than visual assessment. Lung ultrasound, although useful as a bedside adjunct and highly sensitive for subpleural changes, does not capture deep or whole-lung volumetric disease and lacks standardization; therefore, it was not used for primary interstitial lung disease burden quantification in this study. We further concur that Krebs von den Lungen 6 is a promising biomarker for SSc-ILD and future studies on this biomarker, in combination with immune profiling of leukocytes, could be of interest in microbiota research. The presence of a gut–lung axis does not diminish the importance of cardiopulmonary dimensions in SSc-ILD.

In summary, we sincerely appreciate the interest raised in our exploratory investigation. Our international group welcomes participation of other sites and is grateful for the feedback on how to build on this research in our future studies.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25667.

Kristofer Andréasson, MD, PhD
Lund University
Lund, Sweden
Jennifer S. Labus, PhD
Arissa Young, MD
Swapna Joshi, PhD
Jonathan P. Jacobs, MD, PhD
Elizabeth R. Volkmann, MD, MS
evolkmann@mednet.ucla.edu
Grace Hyun Kim, PhD
Jonathan Goldin, MD, PhD
David Geffen School of Medicine, University of California, Los Angeles

1. Andréasson K, Young A, Joshi S, et al. International investigation of the gut–lung axis in systemic sclerosis–interstitial lung disease. Arthritis Care Res (Hoboken) 2026;78(3):352–361.
2. Muller E, Shiryan I, Borenstein E. Multi-omic integration of microbiome data for identifying disease-associated modules. Nat Commun 2024; 15(1):2621.
3. Kim HG, Tashkin DP, Clements PJ, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 2010;28(5 suppl 62): S26–S35.

3. de Pinto M, Coppi F, Spinella A, et al. The predictive role of the TAPSE/sPAP ratio for cardiovascular events and mortality in systemic sclerosis with pulmonary hypertension. Front Cardiovasc Med 2024;11:1430903.
4. Padilla CM, Lafyatis R, Gibson KF, et al. Monocyte LOXHD1 and RHOB expression predictive of progressive systemic sclerosis–associated interstitial lung disease. Arthritis Care Res (Hoboken) 2026;78(4): 456–468.
5. Coppi F, Pagnoni G, Campani C, et al. Sjögren disease and pulmonary hypertension: exploring the intricate link with interstitial lung disease. Int J Cardiol 2025;430:133185.
6. Fairchild RM, Mar DA, Deluna MD, et al. Validation of lung ultrasound interpretation criteria for interstitial lung disease in systemic sclerosis and inflammatory myopathy. Arthritis Care Res (Hoboken) 2025; 77(11):138–1392.

DOI 10.1002/acr.25667

Reply

To the Editor:

We are grateful for the interest from Pagnoni et al in our study.[1] Indeed, this exploratory investigation is, to our knowledge, the first multicenter study linking bacterial species to systemic sclerosis–related interstitial lung disease (SSc-ILD). In addition, it is the first study in the field of systemic sclerosis to adjust for multiple known confounders in analyses of the intestinal microbiota.

We acknowledge the methodologic issues raised by Pagnoni et al, including suggestions on how to probe generalizability across study centers, how to synchronize biospecimen collection, and statistical methodology. Great technical advancements have been made during the last decade in this field, and we appreciate these comments for our future work. We also note other novel approaches on how to integrate whole genome shotgun sequencing with metabolomics and clinical data.[2]

We recognize the concern from Pagnoni et al regarding multiple comparisons in our species-level association analyses. Because this is the first study of its kind, we employed a discovery-based approach emphasizing effect size estimation and hypothesis generation as compared to hypothesis testing. In the context of microbiome research, wherein large sample sizes are required to detect effects, especially under conditions of multiple hypothesis testing, the results suggest potential targets for future hypothesis-driven analyses. In addition, Cohen's *d* was provided to facilitate sample size estimations needed to guide future investigations in this emerging area of research.

To deal with generalizability of effects, we specified a general linear model with a random effect for sites. This model accounts for the variability in outcomes due to differences in site. The estimate for groups is the weighted average of the within-site and between site components of the group differences. Although this model provides the statistical basis for generalization of results across sites, we agree that the leave-one-out sensitivity analyses is an important tool for assessing the influence of each site on the results and the robustness of statistical models. We acknowledge that there may be important site-specific differences or associations with microbiomes because of diversity of diets and environmental exposures between international sites that the random effects model does not capture. Ultimately, replication in well-powered independent datasets is required to confirm the validity and generalizability of the findings.

In agreement with Pagnoni et al, we welcome future integrative "multiaxis" approaches, possibly incorporating disease phenotyping based on circulating biomarkers, immune profiling of leukocytes, and echocardiological assessments. In our study, we used quantitative image analysis of high-resolution computed tomography,[3] a valid approach for approximating the burden of pulmonary fibrosis that is more sensitive than visual assessment. Lung ultrasound, although useful as a bedside adjunct and highly sensitive for subpleural changes, does not capture deep or whole-lung volumetric disease and lacks standardization; therefore, it was not used for primary interstitial lung disease burden quantification in this study. We further concur that Krebs von den Lungen 6 is a promising biomarker for SSc-ILD and future studies on this biomarker, in combination with immune profiling of leukocytes, could be of interest in microbiota research. The presence of a gut–lung axis does not diminish the importance of cardiopulmonary dimensions in SSc-ILD.

In summary, we sincerely appreciate the interest raised in our exploratory investigation. Our international group welcomes participation of other sites and is grateful for the feedback on how to build on this research in our future studies.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25667.

Kristofer Andréasson, MD, PhD
Lund University
Lund, Sweden
Jennifer S. Labus, PhD
Arissa Young, MD
Swapna Joshi, PhD
Jonathan P. Jacobs, MD, PhD
Elizabeth R. Volkmann, MD, MS
evolkmann@mednet.ucla.edu
Grace Hyun Kim, PhD
Jonathan Goldin, MD, PhD
David Geffen School of Medicine, University of California, Los Angeles

1. Andréasson K, Young A, Joshi S, et al. International investigation of the gut–lung axis in systemic sclerosis–interstitial lung disease. Arthritis Care Res (Hoboken) 2026;78(3):352–361.
2. Muller E, Shiryan I, Borenstein E. Multi-omic integration of microbiome data for identifying disease-associated modules. Nat Commun 2024; 15(1):2621.
3. Kim HG, Tashkin DP, Clements PJ, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. Clin Exp Rheumatol 2010;28(5 suppl 62): S26–S35.

DOI 10.1002/acr.25663

Overestimating the risk of febuxostat's hepatotoxicity? comment on the article by Sun et al

To the Editor:

I read with interest the study by Sun et al, which compared hepatotoxicity between febuxostat and benzbromarone in a real-world gout cohort using propensity score matching (PSM).[1] The effort to provide longitudinal comparative data on two commonly used urate-lowering therapies is valuable, particularly given the ongoing debate about hepatic safety. There are some inherent limitations that the authors pointed out, but I would like to highlight a few more factors that are impeding the proper interpretation of results.

First, the PSM algorithm matched only on age, sex, and baseline alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels. After matching, the characteristic of renal disease remained highly imbalanced (24.4% of study participants taking febuxostat vs 6.4% of study participants taking benzbromarone; standardized mean difference [SMD], 0.514) and estimated glomerular filtration rate also remained unbalanced (study participants taking febuxostat, 89.75 mL/min/1.73m^2 vs study participants taking benzbromarone, 94.65 mL/min/1.73m^2; SMD, 0.248).[1]

Renal dysfunction is associated with metabolic syndrome and fatty liver disease, which can predispose a patient to liver enzyme abnormalities.[2,3] So, the increased hepatotoxicity observed with febuxostat treatment could be contributed by this confounding factor rather than a true effect of the drug. Although the authors attempted to negate this by using inverse probability of treatment weighting, residual confounding is highly likely given the magnitude of baseline imbalance. Stratified analyses by renal function would strengthen validity.

Second, in the study, hepatotoxicity was defined as ALT or AST levels greater than three times the upper limit of normal. Most patients experienced grade 1 elevations according to the Common Terminology Criteria for Adverse Events elevations, with no grade 4 events.[4] Importantly, the analysis did not include bilirubin, international normalized ratio, or albumin, thereby precluding evaluation against Hy's law, the accepted predictor of serious drug-induced liver injury (DILI).[5–7] Regulatory guidance emphasizes that ALT or AST elevations alone are poor surrogates for clinically important hepatotoxicity.[5] The current outcome definition may therefore overestimate clinical risk, capturing transient and reversible enzyme abnormalities rather than true DILI.

Third, the study concludes that benzbromarone is safer than febuxostat. Although mild liver function test (LFT) abnormalities appeared less common, this does not address the well-documented concern of rare, fulminant benzbromarone-associated liver failure.[8–10] These rare but severe events drove the withdrawal of benzbromarone from many markets, whereas febuxostat treatment is more often associated with frequent but reversible enzyme elevations.[11,12] Thus, the interpretation based solely on mildly elevated AST or ALT levels can mislead practicing clinicians.

Additionally, the cohort was entirely Chinese, limiting generalizability. Genetics could be one of the confounding factors. Inclusion required at least three LFTs per year, which selected for patients under closer monitoring and introducing potential bias. The study also did not examine dose–response relationships or adherence verification. Collectively, these factors temper the strength of the findings.

I commend the authors for addressing a clinically relevant question. Nevertheless, given the renal parameter imbalance and limited outcome definition, the conclusion that febuxostat carries a greater hepatotoxicity risk than benzbromarone should be interpreted cautiously. I recommend re-analysis stratified by renal function, incorporation of bilirubin-based criteria, and replication in diverse populations before revising clinical practice. There is no direct clinical or pharmacological intervention pertaining to this letter, so ethical clearance was not sought for this letter.

While drafting this letter, I used a large language model (ChatGPT) to help me with the grammar and correction of sentences because English is not my native language.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25663.

Sree Hari Reddy Gadekallu, MD
sreehari.dr@gmail.com
Kurnool Rheumatology Center
Kurnool, India

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score–matched cohort study. Arthritis Care Res (Hoboken) 2025; 77(9):1149–1156.
2. Kuo CF, Yu KH, Luo SF, et al. Gout and risk of non-alcoholic fatty liver disease. Scand J Rheumatol 2010;39(6):466–471.
3. Lee JS, Won J, Kwon OC, et al. Hepatic safety of febuxostat compared with allopurinol in gout patients with fatty liver disease. J Rheumatol 2019;46(5):527–531.
4. *Common Terminology Criteria for Adverse Events (CTCAE)*. Version 5. US Department of Health and Human Services, National Institutes of Health, National Cancer Institute; 2017. Accessed September 6, 2025. https://dctd.cancer.gov/research/ctep-trials/trial-development#ctcae-and-ctep-codes
5. *Guidance for Industry: Drug-Induced Liver Injury: Premarketing Clinical Evaluation*. US Department of Health and Human Services, US Food and Drug Administration; 2009. Accessed September 7, 2025. https://www.fda.gov/media/116737/download
6. European Association for the Study of the Liver. EASL clinical practice guidelines: drug-induced liver injury. J Hepatol 2019;70(6):1222–1261.
7. Temple R. Hy's law: predicting serious hepatotoxicity. Pharmacoepidemiol Drug Saf 2006;15(4):241–243.

8. Zhang MY, Niu JQ, Wen XY, et al. Liver failure associated with benzbromarone: a case report and review. World J Clin Cases 2019; 7(13):1717–1725.
9. Lee MHH, Graham GG, Williams KM, et al. Benefit–risk assessment of benzbromarone in gout. Drug Saf 2008;31(8):643–665.
10. Roberts RL, Wallace MC, Wright DF, et al. Frequency of CYP2C9 polymorphisms in Polynesians and relevance to benzbromarone hepatotoxicity. Joint Bone Spine 2014;81(2): 160–163.
11. Febuxostat. In: LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. National Institute of Diabetes and Digestive and Kidney Diseases; 2012. Accessed September 7, 2025. https://www.ncbi.nlm.nih.gov/books/NBK548645
12. Benzbromarone. In: LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. National Institute of Diabetes and Digestive and Kidney Diseases; 2012. Accessed September 7, 2025. https://www.ncbi.nlm.nih.gov/books/NBK548732

DOI 10.1002/acr.25664

Reply

To the Editor:

Thanks to Dr Gadekallu for raising these critical points regarding hepatic effects of benzbromarone and febuxostat in our study.[1] In that study, stratified analyses based on renal function were performed. However, within the matched cohort, which included a total of 150 patients with renal impairment, hepatotoxicity events occurred in only six of 119 patients in the febuxostat group, whereas no events were observed in the benzbromarone group. The resulting hazard ratio (HR) was reported as an "Inf" value and is therefore not reported. Notably, the majority of patients in the cohort did not have renal disease. Overall, febuxostat treatment was associated with a significantly increased risk of hepatotoxicity compared with benzbromarone (HR, 2.50; 95% confidence interval, 1.13–5.52), a finding consistent with both the primary analysis and the inverse probability of treatment weighting results. The potential influence of renal function on liver enzyme abnormalities warrants further investigation, specifically in a population with significant renal impairment.

The concerns raised regarding the use of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) levels at least three times the upper limit of normal as the sole definition of hepatotoxicity in the study are valid and important. We acknowledge that the definition based solely on ALT or AST levels is a significant limitation. By not measuring bilirubin, international normalized ratio, or albumin, we could not assess for Hy's law.[2] Consequently, our definition likely captured a broader spectrum of liver enzyme abnormalities, many of which may be transient, asymptomatic, and reversible, thereby potentially overestimating the true risk of severe, clinically important drug-induced liver injury (DILI). Our findings should therefore be interpreted as identifying a signal of potential liver injury. Future studies should strive to incorporate bilirubin levels when possible to allow for a more specific risk assessment focused on DILI.

The concern regarding the study's conclusion that benzbromarone is safer than febuxostat represents a misinterpretation. Although our findings indicate that mild liver function test abnormalities were less frequent with benzbromarone, this does not negate the potential risk of severe, fulminant liver failure, a well-documented adverse effect associated with benzbromarone that has prompted its withdrawal in many markets. Such rare yet life-threatening events were not captured in our analysis. Therefore, interpreting the safety profile based on our study may be focused on transient and mild enzyme elevations that are typically reversible. Although febuxostat is associated with more frequent but often transient enzyme elevations, the rare yet severe hepatotoxicity of benzbromarone remains a critical consideration in any risk–benefit assessment.

As noted, the current study has limitations in addressing population-specific factors, dose-dependent risks, and confounding variables, all of which require more rigorously designed research to clarify the hepatic safety profile of urate-lowering drugs in patients with gout. Future studies should include larger sample sizes and longer follow-up periods, with particular emphasis on elucidating the mechanisms and identifying potential biomarkers of severe hepatotoxicity associated with urate-lowering drugs.[3] At the same time, clinical practice should incorporate individualized risk assessment based on factors such as baseline liver and kidney function and concomitant medications, along with enhanced monitoring and patient education to facilitate early detection and management of adverse effects.

Supported by the Shandong Provincial Natural Science Foundation (grant ZR2023QH147 and ZR2023MH213). ChatGPT was used to help with grammar and correction of sentences.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25664.

Wenyan Sun, PhD
Lingling Cui, PhD
Changgui Li, MD
changguili@vip.163.com
Shandong Provincial Clinical Research Center for Immune Diseases and Gout, Shandong Provincial Clinical Medical Center for Gout, The Affiliated Hospital of Qingdao University Qingdao, China

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score-matched cohort study. Arthritis Care Res (Hoboken) 2025;77(9): 1146–1156.
2. Temple R. Hy's law: predicting serious hepatotoxicity. Pharmacoepidemiol Drug Saf 2006;15(4):241–243.
3. Terkeltaub R. Emerging urate-lowering drugs and pharmacologic treatment strategies for gout: a narrative review. Drugs 2023;83(16):1501–1521.

8. Zhang MY, Niu JQ, Wen XY, et al. Liver failure associated with benzbromarone: a case report and review. World J Clin Cases 2019; 7(13):1717–1725.
9. Lee MHH, Graham GG, Williams KM, et al. Benefit–risk assessment of benzbromarone in gout. Drug Saf 2008;31(8):643–665.
10. Roberts RL, Wallace MC, Wright DF, et al. Frequency of CYP2C9 polymorphisms in Polynesians and relevance to benzbromarone hepatotoxicity. Joint Bone Spine 2014;81(2): 160–163.
11. Febuxostat. In: LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. National Institute of Diabetes and Digestive and Kidney Diseases; 2012. Accessed September 7, 2025. https://www.ncbi.nlm.nih.gov/books/NBK548645
12. Benzbromarone. In: LiverTox: Clinical and Research Information on Drug-Induced Liver Injury [Internet]. National Institute of Diabetes and Digestive and Kidney Diseases; 2012. Accessed September 7, 2025. https://www.ncbi.nlm.nih.gov/books/NBK548732

DOI 10.1002/acr.25664

Reply

To the Editor:

Thanks to Dr Gadekallu for raising these critical points regarding hepatic effects of benzbromarone and febuxostat in our study.[1] In that study, stratified analyses based on renal function were performed. However, within the matched cohort, which included a total of 150 patients with renal impairment, hepatotoxicity events occurred in only six of 119 patients in the febuxostat group, whereas no events were observed in the benzbromarone group. The resulting hazard ratio (HR) was reported as an "Inf" value and is therefore not reported. Notably, the majority of patients in the cohort did not have renal disease. Overall, febuxostat treatment was associated with a significantly increased risk of hepatotoxicity compared with benzbromarone (HR, 2.50; 95% confidence interval, 1.13–5.52), a finding consistent with both the primary analysis and the inverse probability of treatment weighting results. The potential influence of renal function on liver enzyme abnormalities warrants further investigation, specifically in a population with significant renal impairment.

The concerns raised regarding the use of alanine aminotransferase (ALT) or aspartate aminotransferase (AST) levels at least three times the upper limit of normal as the sole definition of hepatotoxicity in the study are valid and important. We acknowledge that the definition based solely on ALT or AST levels is a significant limitation. By not measuring bilirubin, international normalized ratio, or albumin, we could not assess for Hy's law.[2] Consequently, our definition likely captured a broader spectrum of liver enzyme abnormalities, many of which may be transient, asymptomatic, and reversible, thereby potentially overestimating the true risk of severe, clinically important drug-induced liver injury (DILI). Our findings should therefore be interpreted as identifying a signal of potential liver injury. Future studies should strive to incorporate bilirubin levels when possible to allow for a more specific risk assessment focused on DILI.

The concern regarding the study's conclusion that benzbromarone is safer than febuxostat represents a misinterpretation. Although our findings indicate that mild liver function test abnormalities were less frequent with benzbromarone, this does not negate the potential risk of severe, fulminant liver failure, a well-documented adverse effect associated with benzbromarone that has prompted its withdrawal in many markets. Such rare yet life-threatening events were not captured in our analysis. Therefore, interpreting the safety profile based on our study may be focused on transient and mild enzyme elevations that are typically reversible. Although febuxostat is associated with more frequent but often transient enzyme elevations, the rare yet severe hepatotoxicity of benzbromarone remains a critical consideration in any risk–benefit assessment.

As noted, the current study has limitations in addressing population-specific factors, dose-dependent risks, and confounding variables, all of which require more rigorously designed research to clarify the hepatic safety profile of urate-lowering drugs in patients with gout. Future studies should include larger sample sizes and longer follow-up periods, with particular emphasis on elucidating the mechanisms and identifying potential biomarkers of severe hepatotoxicity associated with urate-lowering drugs.[3] At the same time, clinical practice should incorporate individualized risk assessment based on factors such as baseline liver and kidney function and concomitant medications, along with enhanced monitoring and patient education to facilitate early detection and management of adverse effects.

Supported by the Shandong Provincial Natural Science Foundation (grant ZR2023QH147 and ZR2023MH213). ChatGPT was used to help with grammar and correction of sentences.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25664.

Wenyan Sun, PhD
Lingling Cui, PhD
Changgui Li, MD
changguili@vip.163.com
Shandong Provincial Clinical Research Center for Immune Diseases and Gout, Shandong Provincial Clinical Medical Center for Gout, The Affiliated Hospital of Qingdao University Qingdao, China

1. Sun W, Cui L, Terkeltaub R, et al. Risk of hepatotoxicity in patients with gout treated with febuxostat or benzbromarone: a propensity score-matched cohort study. Arthritis Care Res (Hoboken) 2025;77(9): 1146–1156.
2. Temple R. Hy's law: predicting serious hepatotoxicity. Pharmacoepidemiol Drug Saf 2006;15(4):241–243.
3. Terkeltaub R. Emerging urate-lowering drugs and pharmacologic treatment strategies for gout: a narrative review. Drugs 2023;83(16):1501–1521.