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. 9 September 2026

Arthritis Care & Research

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

Co-Editors
Kelli D. Allen, PhD
University of North Carolina at Chapel Hill
Durham VA Medical Center
Daniel K. White, PT, ScD, MSc
University of Delaware

Deputy Editors
Joshua Baker, MD, MSCE, *Philadelphia*
Bonnie L. Bermas, MD, *Dallas*
Yvonne C. Lee, MD, MMSc, *Chicago*
Todd A. Schwartz, DrPH, *Chapel Hill*

Editor-in-Chief Emeritus
Marian Hannan, PhD, Boston
Patti Katz, PhD, San Francisco

Journal Publications Committee
Betty Tsao, PhD, *Chair, Charleston*
Elana Bernstein, MD, MSc, *New York*
Krati Chauhan, MD, MPH, *Springfield*
Cynthia Crowson, PhD, *Stewartville*
Himanshu Vashistha, PhD, MBA, *Great Neck*
Marwin Groener, MD, *Baltimore*
Mark Hwang, MD, MS, *Houston*
Eric Roberts, MPH, PhD, *San Francisco*

Associate Editors
Lorinda Chung, MD, MS, *Stanford*
Maureen Dubreuil, MD, *Boston*
Bryant England, MD, *Omaha*
Candace Feldman, MD, MPH, ScD, *Boston*
Angelo Gaffo, MD, MSPH, *Birmingham*
Michael George, MD, MSCE, *Philadelphia*
Afton L. Hassett, PsyD, *Ann Arbor*
April Jorge, MD, *Boston*
Crystal MacKay, PhD, MHSc, BScPT, *Toronto*
Natalie McCormick, PhD, *Boston and Vancouver*
Eli M. Miloslavsky, MD, *Boston*
Susan Murphy, ScD, OT, *Ann Arbor*
Naomi J. Patel, MD, MPH, *Boston*
Julia Simard, ScD, *Stanford*
Louise Thoma, PT, PhD, *Chapel Hill*
Jie Wei, PhD, *Changsha*
Michael H. Weisman, MD, *Palo Alto and Los Angeles*
Pamela F. Weiss, MD, MSCE, *Philadelphia*

Editorial Staff
Susan Case, *Vice President, Strategic Marketing, Communications and Publishing, Maryland*
Maggie Parry, *Director, Quality and Production, Georgia*
Brian Robinson, *Director, Digital Content, Pennsylvania*
Christy Austin, *Publishing Coordinator, Washington*
Kylie Bade, *Managing Editor, North Carolina*

Editorial Board
Graciela S. Alarcón, MD, MPH, *Birmingham*
Carolina Alvarez, MS, *Chapel Hill*
Allen Anandarajah, MBBS, MS, *Rochester*
Liubov Arbeeva, MS, *Chapel Hill*
Shervin Assassi, MD, MS, *Houston*
Ram Bajpai, PhD, *Newcastel-under-Lyme*
Matthew Baker, MD, MS, *Stanford*
Claire Barber, MD, PhD, FRCPC, *Calgary*
Christie Bartels, MD, *Madison*
Jennifer Barton, MD, *Portland*
Teresa J. Brady, PhD, *Atlanta*
Leigh Callahan, PhD, *Chapel Hill*
Laura Cappelli, MD, MHS, MS, *Baltimore*
Rebecca Cleveland, PhD, *Chapel Hill*
Megan E.B. Clowse, MD, MPH, *Durham*
Jamie E. Collins, PhD, *Boston*
Delphine Courvoisier, PhD, *Geneva*
Cynthia Crowson, PhD, *Rochester*
Kimberly DeQuattro, MD, MM, *Stanford*
Cristina Drenkard, MD, PhD, *Atlanta*
Jeffrey Driban, PhD, AT Ret, CSCS, *Boston*
Alyssa B. Dufour, PhD, *Boston*
Amanda Eudy, PhD, *Durham*
Robert Fairchild, MD, PhD, *Stanford*
Titilola Falasinnu, PhD, *Stanford*
Elizabeth Ferucci, MD, MPH, *Anchorage*
Tracy Frech, MD, MS, *Salt Lake City*
Shivani Garg, MD, PhD, *New Haven*
Yvonne Golightly, PhD, *Chapel Hill*
Jaime Guzman, MD, MSc, *Vancouver*
Michelle Hall, PhD, MSc, BSc(Hons), *Sydney*
Lauren Henderson, MD, MMSc, *Boston*
Monique Hinchcliff, MD, MS, *New Haven*
Joel Hirsh, MD, *Denver*
Dan Horton, MD, MSCE, FISPE, *New Brunswick*
Mark Hwang, MD, MS, *Houston*
Tate Johnson, MD, *Omaha*
Guy Katz, MD, *Boston*
Deepak Kumar, PT, PhD, *Boston*
Linda Li, PhD, *Vancouver*
Shao-Hsien Liu, PhD, *Worchester*
Grace Hsaio-Wei Lo, MD, MSc, *Houston*
Bing Lu, PhD, *Farmington*
Sara McCoy, MD, PhD, *Madison*
Christopher Mecoli, MD, MHS, *Baltimore*
Devyani Misra, MD, MS, *Boston*
Gulsen Ozen, MD, *Omaha*
Elizabeth Park, MD, MSc, *New York*
Sofia Pedro, MSc, *Wichita*
Anthony Perruccio, PhD, *Toronto*
Jennifer Pierce, PhD, *Ann Arbor*
Amir Razmjou, MD, MSc, *Los Angeles*
Joshua Stefanik, MSPT, PhD, *Boston*
Sara Tedeschi, MD, MPH, *Boston*
Lauren Terhorst, PhD, *Piitsburgh*
Marinka Twilt, MD, MScE, *Calgary*
Manuel Ugarte-Gil, MD, MSc, *Lima*
Antonia Valenzuela, MD, MS, *Santiago*
Ernest Vina, MD, MS, *Tucson*
Elizabeth Volkmann, MD, MS, *Los Angeles*
Zachary Wallace, MD, MSc, *Boston*
Katie Wysham, MD, *Seattle*
Chao Zeng, MD, PhD, *Changsa*
Guohao (Arthur) Zhu, PhD, *Ann Arbor*

Association of Rheumatology Professionals
2025–2026 Executive Committee
Janell Martin, CAE, Executive Director, *Atlanta*

Becki Cleveland, PhD, *Chapel Hill*, President
Kaleb Michaud, PhD, *Omaha*, President-Elect
Adam Goode, DPT, PhD, PT, *Durham*, Immediate Past President
Kimberly Steinbarger, DHSc, MHS, PT, *Bangor*, Secretary
Christine Pellegrini, PhD, *Columbia*, Committee on Research Liaison
Victoria Merrell, PA-C, MPT, *Encinitas*, Committee on Government Affairs Liaison
Shelby Jones, APRN, MSN, NP, *Houston*, eLearning Subcommittee Chair
Katharine McCarthy, PharmD, *Rochester*, Interprofessional Practice & Management Chair
Charmayne Dunlop-Thomas, MA, MS, *Atlanta*, Convergence Planning Subcommittee Chair
Talitha Black, MA, OT, *Los Angeles*, Member-at-Large
Annelle Reed, CPNP, MSN, *Birmingham*, Member-at-Large Finance
Stephen Moore, PhD, *Birmingham*, Member-at-Large
Anne Bass, MD, FACR, *New York*, ACR Invited Guest

Copyright © 2026 American College of Rheumatology. All rights reserved, including rights for text and data mining and training of artificial technologies or similar technologies. No part of this publication may be reproduced, stored or transmitted in any form or by any means without the prior permission in writing from the copyright holder. Authorization to copy items for internal and personal use is granted by the copyright holder for libraries and other users registered with their local Reproduction Rights Organization (RRO), e.g. Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923, USA (www.copyright.com), provided the appropriate fee is paid directly to the RRO. This consent does not extend to other kindsof copying such as copying for general distribution, for advertising or promotional purposes, for creating new collective works or for resale. Special requests should be addressed to: permissions@wiley.com

Access Policy: Subject to restrictions on certain backfiles, access to the online version of this issue is available to all registered Wiley InterScience users 12 months after publication. Subscribers and eligible users at subscribing institutions have immediate access in accordance with the relevant subscription type. Please go to onlinelibrary.wiley.com for details.

The views and recommendations expressed in articles, letters, and other communications published in *Arthritis Care & Research* are those of the authors and do not necessarily reflect the opinions of the editors, publisher, or American College of Rheumatology. The publisher and the American College of Rheumatology do not investigate the information contained in the classified advertisements in this journal and assume no responsibility concerning them. Further, the publisher and the American College of Rheumatology do not guarantee,warrant, or endorse any product or service advertised in this journal.

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 • SEPTEMBER 2026 • NO. 9

Special Article

Clinicopathologic Conference: The Plot Thickens: From a Classic Presentation to Bilateral Renal Lesions
Frank Cai, Kirtan Patolia, Rebecca D. Sullenger, David Howell, Shih-Hsiu J. Wang, Bangchen Wang, Megan E. B. Clowse, Lena Eder, and Jeffrey Shen 1089

Editorial: Common Data Elements in Autoimmune Disease Research
Victoria K. Shanmugam, Carmen Ufret-Vincenty, Xinrui Li, Anne Deslattes-Mays, Richard H. Scheuermann, Belinda Seto, and Susan Gregurick 1098

Myositis

Standardized Interoperable Data Collection for Myositis Research: Developing Expert Consensus on Common Data Elements for Myositis Outcome Measures
Didem Saygin, Matthew Diller, Varsha Surampudi, Mark Bodkin, Payam Noroozi Farhadi, Christopher A. Mecoli, Audrey Kessel, Rohit Aggarwal, Helene Alexanderson, Anthony Amato, Christie M. Bartels, Olivier Benveniste, Michelle Best, Hector Chinoy, Ingrid de Groot, Brian Feldman, Adam M. Huber, Hanna Kim, Susan Kim, Linda Kobert, Valerie Leclair, Manuel Lubinus, Pedro M. Machado, Andrew Mammen, Liza J. McCann, Tahseen Mozaffar, Chester Oddis, Julie J. Paik, Angelo Ravelli, Nicolino Ruperto, Jens Schmidt, Ellen Werner, Victoria Werth, Adam Schiffenbauer, Richard H. Scheuermann, and Lisa G. Rider, for the IMACS Myositis CDE Working Group 1101

Marked Long-Term Improvement in Lung Function in Melanoma Differentiation–Associated Protein 5 Antibody–Positive Dermatomyositis Patients: Experience of a Single-Center Longitudinal Cohort in North America
Jenice X. Cheah, Sangmee S. Bae, Tiffany De Leon, Yuna Lee, Rong Guo, David Elashoff, Jennifer Wang, Ani Shahbazian, and Christina Charles-Schoeman 1112

Immune Checkpoint Inhibitor–Related Myositis and Associated Triad Overlap Syndrome
Selene Rubino, Grant W. Cannon, Brian C. Sauer, Jorge Rojas, Gregory J. Stoddard, Gary A. Kunkel, Jessica A. Walsh, Bryant R. England, Joshua F. Baker, and Tawnie J. Braaten 1123

Systemic Sclerosis

Mycophenolate Mofetil Treatment Reduces the Risk of Treatment Escalation Due to Vascular Complications in Limited Cutaneous Systemic Sclerosis: Emulation of a Target Trial From the Italian Rheumatology Society SPRING Registry
Enrico De Lorenzis, Gerlando Natalello, Rossella De Angelis, Lucrezia Verardi, Dilia Giuggioli, Gianluigi Bajocchi, Lorenzo Dagna, Silvia Bellando-Randone, Giovanni Zanframundo, Rosario Foti, Fabio Cacciapaglia, Giovanna Cuomo, Alarico Ariani, Edoardo Rosato, Gemma Lepri, Francesco Girelli, Valeria Riccieri, Elisabetta Zanatta, Ilaria Cavazzana, Francesca Ingegnoli, Maria De Santis, Giuseppe Murdaca, Giuseppina Abignano, Giorgio Pettiti, Alessandra Della Rossa, Maurizio Caminiti, Annamaria Iuliano, Giovanni Ciano, Lorenzo Beretta, Gianluca Bagnato, Ennio Lubrano, Maria Ilenia De Andres, Alessandro Giollo, Cosimo Bruni, Martina Orlandi, Marco Fornaro, Marta Saracco, Cecilia Agnes, Pier Giacomo Cerasuolo, Gabriella Alonzi, Edoardo Cipolletta, Federica Lumetti, Amelia Spinella, Luca Magnani, Corrado Campochiaro, Giacomo De Luca, Veronica Codullo, Elisa Visalli, Carlo Iandoli, Antonietta Gigante, Greta Pellegrino, Erika Pigatto, Maria-Grazia Lazzaroni, Franco Franceschini, Elena Generali, Gianna Mennillo, Simone Barsotti, Giuseppa Pagano Mariano, Federica Furini, Licia Vultaggio, Simone Parisi, Clara Lisa Peroni, Gerolamo Bianchi, Enrico Fusaro, Gian Domenico Sebastiani, Marcello Govoni, Salvatore D'Angelo, Franco Cozzi, Fabrizio Conti, Serena Guiducci, Andrea Doria, Carlo Salvarani, Florenzo Iannone, Maria Antonietta D'Agostino, Clodoveo Ferri, Marco Matucci Cerinic, and Silvia Laura Bosello, on behalf of the SPRING Italian Registry Collaborators 1130

Differential Item Functioning on the Patient Health Questionnaire 8 by Disease Subtype, Language, Sex, and Age Among People With Systemic Sclerosis: A Scleroderma Patient-Centered Intervention Network Cohort Study
Sophie Hu, Marie-Eve Carrier, Marie-Claude Geoffroy, Meira Golberg, Linda Kwakkenbos, Susan J. Bartlett, Catherine Fortuné, Amy Gietzen, Karen Gottesman, Geneviève Guillot, Laura K. Hummers, Amanda Lawrie-Jones, Vanessa L. Malcarne, Michelle Richard, Maureen Sauvé, Luc Mouthon, Andrea Benedetti, and Brett D. Thombs, on behalf of the SPIN Investigators 1140

Rheumatoid Arthritis

Use of a Molecular Signature Response Classifier to Inform Treatment Selection Improves Clinical Disease Activity Among Patients with Rheumatoid Arthritis Initiating a Biologic or Targeted Synthetic Disease-Modifying Antirheumatic Drug
Fenglong Xie, Timothy Beukelman, Nicholas P. McCormick, and Jeffrey R. Curtis 1149

Impact of Dual Rheumatoid Factor and Anticitrullinated Protein Antibody Seropositive, Single Seropositive, and Seronegative Rheumatoid Arthritis on Outcomes
Rebecca T. Brooks, Sara J. Achenbach, Vanessa L. Kronzer, Elena Myasoedova, Cynthia S. Crowson, and John M. Davis III 1159

Evaluating a Pragmatic Strength Alternative for Frailty Measurement and Assessing Its Predictive Capacity Against Established Frailty Instruments in Rheumatoid Arthritis
Kylie E. Riggles, Hannah F. Brubeck, Adrienne D. Tanus, Courtney N. Loecker, Punyasha Roul, Bryant R. England, Elizabeth R. Wahl, James S. Andrews, Namrata Singh, Joshua F. Baker, Patricia P. Katz, Dolores M. Shoback, Jose M. Garcia, Ariela R. Orkaby, and Katherine D. Wysham 1167

Systemic Lupus Erythematosus

Addressing Economic Insecurities Can Improve Patient-Reported Outcomes in Lupus
Jay Patel, Tripti Singh, Meredith Ingersoll, Shelby Gomez, Amanda Weber, Sarah E. Panzer, Sancia Ferguson, Christie M. Bartels, and Shivani Garg 1177

Brief Report: Systemic Lupus Erythematosus Mortality Among Decedents Aged ≥15 Years—United States, 2018–2023
Danielle Dawson, Kurt J. Greenlund, and Kamil E. Barbour 1188

Brief Report: Rheumatologic Manifestations of Patients With Type B Insulin Resistance
S. Amara Ogbonnaya, Sandra G. Williams, Raphael A. Kirou, Marissa Lightbourne, and Rebecca J. Brown 1195

Osteoarthritis

Review Article: Conflicts of Interest Reporting in Trials and Guidelines Addressing Glucocorticoid Injections for Knee Osteoarthritis
Craig Vecchiarelli, Lee Newman, Shannon T. Boyd, Jodi L. Young, and Daniel I. Rhon 1201

Associations of Sleep and Shift Work With Osteoarthritis Risk
Elizabeth L. Yanik, Abigail Bridgeman, Erik D. Herzog, Vy Pham, Bradley A. Evanoff, and Farshid Guilak 1209

Does Long-Term Lower Extremity Strength Training in Adults With Knee Osteoarthritis and Varus Alignment Reduce Knee Joint Loading During Gait?
Stephen P. Messier, Matthew N. Vigliotti, Paige E. Rice, Brian Pietrosimone, Shannon L. Mihalko, Edward H. Ip, Richard F. Loeser, David J. Hunter, Ali Guermazi, Ryan Hill, Santiago Saldana, Kim L. Bennell, and Paul DeVita 1220

Association of Wearable Sensor–Derived Gait Measures With Cartilage Damage Over Two Years Among Adults With or at Risk of Knee Osteoarthritis
Kathryn L. Bacon, David T. Felson, S. Reza Jafarzadeh, Jeffrey M. Hausdorff, Eran Gazit, Ali Guermazi, Frank W. Roemer, Neil A. Segal, Cora E. Lewis, Michael C. Nevitt, and Deepak Kumar, the Multicenter Osteoarthritis Study Investigators 1229

Moderators of the Effects of Exercise and Manual Therapy for People With Knee and Hip Osteoarthritis: A Secondary Analysis of a Randomized Clinical Trial
Daniel Cury Ribeiro, Sarah E. Lamb, and J. Haxby Abbott 1239

Fibromyalgia

New Onset of Fibromyalgia After Exposure to a Combat Environment: A Longitudinal Cohort Study
Jay B. Higgs, Willie J. Hale, Casey L. Straud, Jim Mintz, Stacey Young-McCaughan, Kimberly D. Gomes, Chelsea J. Sterne, Katrina M. Lawrence-Wolff, Alan J. Bartholomew, Kevin M. Kelly, Douglas M. Maurer, Catherine Vriend, Brett T. Litz, Douglas E. Williamson, and Alan L. Peterson, for the STRONG STAR Consortium 1247

Vasculitis

Subcutaneous Versus Intravenous Tocilizumab in Aortitis Associated With Giant Cell Arteritis: Multicenter Study of 196 Patients
Carmen Secada-Gómez, Javier Loricera, Adrián Martín-Gutiérrez, Javier Narváez, Vicente Aldasoro, Olga Maiz, Paloma Vela, Susana Romero-Yuste, Eugenio de Miguel, Eva Galíndez-Agirregoikoa, Jesús C. Fernandez-López, Iván Ferraz-Amaro, Julio Sanchez-Martín, Patricia Moya-Alvarado, Cristina Campos, Santos Castañeda, and Ricardo Blanco, on behalf of the Tocilizumab in Giant Cell Arteritis Spanish Collaborative Group 1256

Gout

Brief Report: Effect of Implementing a Dashboard With or Without a Best Practice Alert on HLA-B*58:01 Testing Rates Among Allopurinol Users at VA Medical Centers
Abimbola Fadairo-Azinge, Gabriela Schmajuk, Sristi Sharma, Colleen P. Flanagan, Sara Faghihi-Kashani, Ning Zhang, and Mary A. Whooley 1265

Rheumatology Education & Practice

Brief Report: Variations in the Adult Rheumatology Fellowship Training in the United States: Opportunities and Challenges
Debendra Pattanaik, Roxanna Sabghi, Govind S. Bindra, and Syed Raza 1270

Other and Mixed Rheumatic Diseases

Utilization of Rehabilitation Services Among People With Common Autoimmune Rheumatic Diseases: A Systematic Review
Joel R. Thompson, Keri Geinosky, Josh C. Torrey, Astia Allenzara, and Louise M. Thoma 1277

Major US Formulary Change Triggers, but Does Not Sustain, Increases in Biosimilar Adalimumab Prescribing
Eric T. Roberts, Gabriela Schmajuk, Mackenzie Clark, Diana Ung, and Jinoos Yazdany 1287

Brief Report: Reproductive Health Counseling in Rheumatology: Gaps, Barriers, and Patient Impact
Daniele Marcy, Kristin Sturm, Tara Skorupa, Caroline Maretz, Guiset Carvajal Bedoya, Aaron Lazorwitz, Ju Young Kwag, M. Kristen Demoruelle, and JoAnn Zell 1298

Letter

Within-Day Activity Patterns in Knee Osteoarthritis: Frailty Signal or Treatment Marker? Comment on the Article by Liles et al
Ruoyao Cheng, Caihua He, and Ying Liu 1304

Reply

Daniel K. White and Sydney C. Liles 1305

Cover image: The image on the cover (De Lorenzis et al; Supplemental Figure 2a) shows the geographical distribution of patients with limited cutaneous systemic sclerosis enrolled in the Italian Society of Rheumatology Systemic Sclerosis Progression INvestiGation (SPRING) registry.

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1089–1097
DOI 10.1002/acr.80019
© 2026 American College of Rheumatology.

American College of Rheumatology
Empowering Rheumatology Professionals

CLINICOPATHOLOGIC CONFERENCE

The Plot Thickens: From a Classic Presentation to Bilateral Renal Lesions

Frank Cai, MD,[1] Kirtan Patolia, MD,[2] Rebecca D. Sullenger, MPH,[3] David Howell, MD, PhD,[1] Shih-Hsiu J. Wang, MD, PhD,[1] Bangchen Wang, MD, PhD,[1] Megan E. B. Clowse, MD, MPH,[1] Lena Eder, MD,[1] and Jeffrey Shen, MD[1]

CASE PRESENTATION

History of present illness

The patient, a 63-year-old woman, presented to our hospital with one week of severe bilateral thigh pain, bilateral jaw pain, and a transient episode of left-sided painless vision loss. In the previous four months, she had experienced progressive fatigue, intermittent fevers, night sweats, and generalized weakness. Three months prior, she had presented to her primary care provider with an episode of fever measured at 100.6°F, sinus pressure, and nasal discharge. She was known to have a history of recurrent sinusitis, which was managed with intermittent courses of antibiotics and steroids, with temporary symptom relief. Workup during this episode was significant for a new normocytic anemia (hemoglobin level 10.7 g/dL), thrombocytosis (592 × 10^9/L), and an elevated erythrocyte sedimentation rate (ESR) of 86 mm/hr. Her C-reactive protein (CRP) level and white blood cell count were within normal range. She was again treated for a presumed sinus infection with 10 days of clindamycin and 4 days of prednisone 20 mg daily, with significant improvement in symptoms. However, over the next two weeks, her significant fatigue, night sweats, and generalized weakness recurred. One month before her current presentation, progressive worsening of these symptoms prompted hospital admission, and testing demonstrated worsening anemia (hemoglobin level 7.8 g/dL). Additional workup of the anemia suggested mixed iron deficiency anemia and anemia of chronic disease. Markers of inflammation were rechecked and showed an elevated CRP level (3.47 mg/dL) and ESR (54 mm/hr). Blood culture results were negative, and the transthoracic echocardiogram obtained to evaluate for endocarditis was without evidence of vegetations or clots. The etiology of her symptoms remained unclear, and she was discharged with recommendations for supportive care for ongoing symptoms and plans to continue investigation as an outpatient. An upper endoscopy and colonoscopy were completed shortly afterward to evaluate for causes of blood loss in the setting of iron deficiency anemia, which were unrevealing, and random biopsies that were taken showed no abnormalities.

Over the week preceding her current presentation, she experienced new symptoms, including constant and progressive bilateral thigh pain. The pain was exacerbated by activity and improved at rest, and its severity significantly limited her functional capacity. She denied having associated morning stiffness or improvement throughout the day with use. She also noticed new bilateral jaw pain, which was present only upon chewing. It was characterized as dull and aching and would resolve immediately after cessation of chewing. Then a transient episode of left-sided painless vision loss occurred the day before presentation and lasted for a few minutes. She stated she could see only a "gray screen," and it subsequently self-resolved without recurrence. She also reported a history of chronic, intermittent right-sided retro-orbital headaches which had been ongoing for many years, occurring approximately every six weeks and lasting less than 24 hours, and were characterized as a dull, achy sensation associated with nausea, vomiting, and photosensitivity. These headaches were relieved with acetaminophen and ibuprofen use, and lying down in a dark room was also helpful. She endorsed triggers such as stress and menstrual periods. Her fatigue, intermittent fevers, night sweats, and generalized weakness remained ongoing symptoms. Although she endorsed that the overall severity of her nasal drainage had been worse in recent months, she denied having any current sinus symptoms at the time of presentation.

[1]Duke University Hospital, Durham, North Carolina; [2]Barnes-Jewish Hospital, St Louis, Missouri; [3]Duke University School of Medicine, Durham, North Carolina.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80019.

Address correspondence via email to Jeffrey Shen, MD, at Jshenrheum@gmail.com.

Submitted for publication October 16, 2025; accepted in revised form February 4, 2026.

Past medical history

The patient had a medical history of chronic sinusitis, migraines, osteoarthritis, and scoliosis. Her comprehensive medication list included fluoxetine, quetiapine as needed, alprazolam as needed, vitamin B_{12} supplements, fluticasone propionate nasal spray, loratadine, oral progesterone, and an estradiol patch.

Social and family history

The patient denied having any current or prior tobacco or illicit drug use but endorsed occasional alcohol use. Her paternal aunt had an early onset of breast cancer at age 30 years, and her father had lung cancer. She denied having any family history of autoimmune disease. She also denied having any recent travel history.

Review of systems

The patient reported bilateral thigh pain, bilateral jaw pain, and an episode of transient left-sided painless vision loss in addition to her fatigue, intermittent fevers, night sweats, generalized weakness, nasal drainage with sinus pressure, and unchanged chronic headaches. She denied having any weight loss, scalp tenderness, or shoulder pain or stiffness. An additional comprehensive review of systems was completed, and results were otherwise negative.

Physical examination

On examination, she was afebrile and hemodynamically stable and had a normal respiratory rate and oxygen saturation. She was otherwise diaphoretic and ill appearing. Tenderness over the scalp and bilateral temporal arteries were absent, but she had minimal bilateral temporomandibular tenderness to palpation with range of motion testing. The cranial nerve examination did not reveal palsies. No conjunctival, episcleral, or scleral erythema or pain with extraocular movements was detected. The ear examination showed a well-formed pinna without lesions, erythema, or edema. Gross hearing was normal. The nasal examination was notable for right-sided maxillary sinus pressure and right septal deviation without perforation. No active nasal discharge, ulcers, turbinate edema or erythema, crusting, or saddle nose deformity was identified. The oral cavity showed moist mucous membranes with normal dentition and no ulcers. The trachea was midline with normal speech. The parotid glands were not enlarged, and no lymphadenopathy was palpable. Examination of the cardiovascular, respiratory, and abdominal systems was unremarkable. No synovitis, effusions, joint tenderness, or restrictions in range of motion were identified. Motor strength examination noted bilateral 4/5 hip flexion and 4/5 knee flexion, limited by pain. Strength was otherwise full. Peripheral pulses were 2+, and no lower extremity edema was detected. She had normal sensation in her upper and lower extremities. No abnormal cutaneous rashes, purpura, petechiae, ulcers, or nodules were detected.

An ophthalmology evaluation with a slit lamp, fundoscopy, and optical coherence tomography did not show optic disc edema, cotton wool spots, retinal artery occlusions, or areas of definitive ocular ischemia. Fluorescein angiography was negative for delayed or absent choroidal and retinal arterial perfusion, occlusive disease of the posterior ciliary arteries, and areas of retinal or optic nerve ischemia.

Laboratory evaluation

On the current admission, the patient's complete blood count was notable for an elevated white blood cell count (15.3×10^9/L), with neutrophilic predominance (12.4×10^9/L), normal lymphocyte and eosinophil counts, normocytic anemia (hemoglobin level 8.6 g/dL), and thrombocytosis (690×10^9/L). The peripheral smear with flow cytometry showed leukocytosis with neutrophilia and monocytosis, normocytic anemia and thrombocytosis, and no immunophenotypically abnormal leukocyte or expanded immature population. Markers of inflammation were elevated (ESR 56 mm/hr, CRP level 17.86 mg/dL). The creatine kinase (CK) level was within normal range. The admission creatinine level was 0.7 mg/dL. Urinalysis was unremarkable, without protein, glucose, blood, red blood cells, white blood cells, or abnormal casts. A spot urine protein test result was negative. Results of infectious tests in the serum for syphilis, HIV, Epstein-Barr virus, hepatitis B virus, hepatitis C virus, Lyme disease, and tuberculosis were negative. Results of a respiratory virus polymerase chain reaction panel for influenza, respiratory syncytial virus, and SARS-CoV-2 were also negative. Two sets of peripheral blood cultures did not show growth after 72 hours. Serum protein electrophoresis (SPEP) showed mild asymmetry in the beta and gamma regions, and follow-up immunofixation did not identify a monoclonal component. Urine protein electrophoresis (UPEP) was also negative for a paraprotein peak. Ig levels, including IgG, were normal. Antinuclear antibody and antiphospholipid antibody test results were also negative. Then, three days into the admission, results of antineutrophil cytoplasmic antibody (ANCA) studies were a cytoplasmic ANCA (c-ANCA) titer of 1:160, an anti–proteinase 3 (anti-PR3) antibody level of >8.0U/mL (reference range ≤1.0U/mL), and negative antimyeloperoxidase (anti-MPO) antibodies.

Concurrently on admission, a computed tomography angiogram (CTA) of the head and neck was obtained and did not show evidence of an acute intracranial abnormality, intracranial large vessel occlusion, or a hemodynamically significant arterial stenosis in the neck. The day after, a bilateral temporal artery ultrasound was normal, without a halo sign suggestive of active arteritis, and all vessels were compressible bilaterally. Three days into her admission, a CTA of the chest, abdomen, and pelvis

showed a few <5-mm solid pulmonary nodules and a widely patent thoracic and abdominal vasculature without evidence of stenosis, wall thickening, or fat stranding. Notably, it demonstrated persistent asymmetric cortical hypoenhancement of the bilateral kidneys, which was noted to also be present on a CTA of the abdomen and pelvis obtained one month prior as part of her workup (Figure 1). No further diagnostic or treatment interventions for these findings have been previously pursued. There was no lymphadenopathy or hepatosplenomegaly. A temporal artery biopsy (TAB) and renal biopsy were subsequently obtained.

DIFFERENTIAL DIAGNOSIS

The presenting symptoms, elevated markers of inflammation, and bilateral renal hypodensities visualized on CTA were concerning for a systemic, multisystem process. The jaw pain that worsened with chewing and improved with rest was concerning for claudication, but a musculoskeletal etiology was also considered. The patient experienced transient monocular painless vision loss that she described as a "gray screen," sharing similarities with the description of a curtain covering an eye, which was concerning for amaurosis fugax. Although other common headache syndromes, such as migraines, can also cause transient vision loss, amaurosis fugax needed to be evaluated first given the fact that it is a not-to-miss emergency. Amaurosis fugax is typically caused by vascular compromise, and therefore the concern for an underlying vasculitis or vasculopathy was raised. Although the chronic headache was considered as an associated manifestation, it was favored to be secondary to migraines given the characteristic description and chronic stability with episodes independent of her new ongoing symptoms and therefore was deemphasized during the differential consideration. Given the lack of evolution of the bilateral renal lesions on CTA compared to one month prior, pyelonephritis and renal infarcts were considered less likely, and instead, consideration for infiltrative diseases was invoked. A broad differential including infections, neoplastic disorders, infiltrative diseases, and systemic vasculitides was considered.

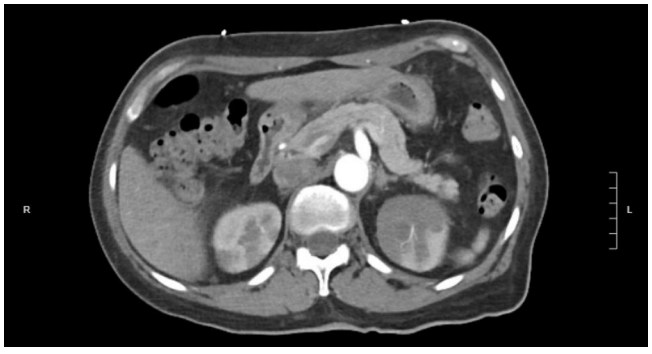

Figure 1. Computed tomography angiogram of the abdomen and pelvis showing asymmetric left-greater-than-right hypoenhancement of the bilateral kidneys.

Infections. Infective endocarditis can cause chronic constitutional symptoms and embolic phenomena that can lead to amaurosis fugax and renal cortical infarcts.[1] However, embolic events are typically persistent rather than the transient vision loss that was described.[2] Myalgias can be seen, but severe pain in the thighs is unusual. Finally, chronic sinusitis is also not a typical symptom, as the sinonasal mucosa is not a target for embolic spread.[3] The prior transthoracic echocardiogram, which showed no evidence of vegetations, and negative blood culture results also make endocarditis less likely. A transesophageal echocardiogram could be pursued if the rest of the workup for other etiologies is unrevealing.

Mycobacterium tuberculosis (TB) is another important chronic infection to consider. Extrapulmonary TB occurs in 10% to 42% of cases and can involve any organ, including the musculoskeletal system, causing myalgias; craniofacial structures, causing jaw pain; the eye, leading to vision loss; and the kidneys, presenting as infiltrative lesions.[4] The infiltrating renal lesions can be asymptomatic. However, urinalysis would usually show asymptomatic hematuria or sterile pyuria.[5] The eye is usually involved as uveitis, but fluctuating vision loss can occur in patients with tuberculous meningitis, though our patient lacked other clinical features of meningitis.[6] TB also usually has pulmonary involvement, which was absent in our patient. The negative interferon-γ release assay result further decreases the likelihood, although the sensitivity of this test in active TB is limited, and tissue acid-fast bacilli staining and culture would be more definitive.[7]

Syphilis was also considered, as it can present with a wide array of systemic manifestations, especially in the secondary and tertiary stages. The eye can be involved at any stage of syphilis, and ischemic retinitis has been described, although this would present with more persistent symptoms, including blurry vision and blind spots.[8] Transient vision loss could also theoretically occur as a transient ischemic attack from meningovascular syphilis, but other associated symptoms were lacking.[9] Fever and myalgias may also occur, and even temporal arteritis has been described with syphilis.[10,11] Syphilis is also known to cause aortitis or endarteritis, which may lead to renal infarction.[12] Our patient had negative serologic testing results for syphilis.

Disseminated fungal infections, particularly aspergillosis and mucormycosis, are important considerations. Both can cause invasive sinusitis, musculoskeletal pains, and renal infiltrative lesions that can cause wedge-shaped infarcts. Supporting imaging features are key to diagnosis, indicating areas of involvement such as the sinuses, which typically will show bony or soft tissue extension that often results in necrosis. This was not present in our patient. She also lacked the usual predisposing risk factors for developing disseminated aspergillosis or mucormycosis, such as an immunocompromised state, neutropenia, or uncontrolled

diabetes mellitus.[13,14] To complete the evaluation, a biopsy of the affected tissue with fungal stains and culture should be pursued.

Neoplastic disorders. Primary renal lymphomas are very rare and can have an infiltrative appearance on CTA.[15] Myeloid neoplasms can also infiltrate the kidneys.[16,17] However, infiltration by these hematologic malignancies is usually accompanied by renal dysfunction and urine abnormalities, including hematuria and proteinuria, which are absent in our patient.[18,19] Lymphomas and leukemias can cause chronic constitutional symptoms, including myalgias, and can cause a presentation mimicking temporal arteritis through direct vascular infiltration or as a vasculitis through a paraneoplastic phenomenon. One subtype to consider is Waldenström's macroglobulinemia (lymphoplasmacytic lymphoma).[20] Hyperviscosity in this condition can lead to headache and transient vision loss, but this is usually binocular and often accompanied by epistaxis and other mucocutaneous bleeds.[21] Rarely, Waldenström macroglobulinemia can infiltrate the central nervous system to cause Bing-Neel syndrome, which can result in jaw pain and vision loss.[22] Deposition of paraprotein in the temporal arteries has also been reported in Bing-Neel syndrome.[23] However, patients usually have lymphadenopathy, hepatosplenomegaly, cytopenias, or peripheral neuropathy as accompanying features. Chronic sinusitis is also not a commonly described manifestation.[20] SPEP/UPEP followed by immunofixation should be obtained, the results of which were not suggestive of hematologic malignancy in our patient. The peripheral smear also did not identify any blasts, atypical lymphocytes, or monomorphic white blood cells. Tissue histopathology is currently pending.

Sarcoidosis. Sarcoidosis is a multisystem granulomatous disease that could explain several key features of our patient's presentation. Chronic disease activity can lead to constitutional symptoms, including fevers. Sarcoid myopathy can cause bilateral thigh pain with normal CK levels, consistent with our patient's findings. Nasal mucosal involvement can also result in chronic sinusitis.[24] Renal sarcoidosis can rarely manifest as infiltrative parenchymal lesions, potentially explaining the bilateral cortical hypoenhancing lesions on CTA imaging. However, sarcoidosis almost always demonstrates characteristic chest imaging findings, particularly bilateral hilar lymphadenopathy or parenchymal lung involvement, which were absent on our patient's CTA. The transient monocular vision loss in our patient suggests vascular ischemia rather than the infiltrative or chronic inflammatory processes typical of ocular involvement from sarcoidosis. Most importantly, renal involvement in sarcoidosis is typically accompanied by hypercalcemia, impaired renal function, or urinary abnormalities—none of which were present in our patient.[25] Although isolated extrapulmonary sarcoidosis can occur, it represents only a small minority of cases.

IgG4-related disease. IgG4-related disease (IgG4-RD) is a chronic multisystem inflammatory disease that can present with mass-like renal lesions or infiltrate the kidneys, causing the CTA appearance seen in our patient.[26] Chronic sinusitis is also a commonly associated allergic feature. Ocular and neurologic involvement is also well described, although this condition typically causes a more progressive orbital inflammation that leads to persistent vision loss rather than a transient episode through extraocular compression. Notably, IgG4-RD is increasingly recognized as a variable vessel vasculitis that could cause aortitis or temporal arteritis presenting as amaurosis fugax, which could present as transient vision loss.[27,28] Fever, constitutional symptoms, and myalgias may rarely occur when IgG4-RD causes vasculitis but otherwise are atypical manifestations. Although serum IgG4 concentration elevations should not be considered specific for diagnosis, they can be supportive. Our patient had normal Ig levels, including IgG, though IgG subclasses were not obtained. The patient also lacked other major manifestations of the disease, including autoimmune pancreatitis, tubulointerstitial nephritis, salivary gland infiltration, or other evidence of organ fibrosis. The presentation can be heterogeneous, and histopathology including storiform fibrosis, obliterative phlebitis, and characteristic lymphoplasmacytic infiltration is supportive.[29]

Amyloidosis. Amyloidosis causes chronic multisystem disease and notably can affect the temporal arteries through deposition of amyloid proteins, leading to vascular lumen narrowing as well as possibly serving as a trigger for a foreign-body inflammatory reaction. This is well described in AL amyloidosis, which can cause symptoms of jaw claudication and vision loss.[30] Although muscle involvement is rare, when it occurs, it can cause both weakness and myalgias with normal CK levels.[31] Chronic sinusitis is also rare but has been reported in localized disease affecting the sinonasal tract.[32] Infiltration into the kidneys may mimic renal infarcts on imaging.[33,34] AL amyloidosis can also occur in association with monoclonal gammopathy of unknown significance, multiple myeloma, or Waldenström macroglobulinemia.[35] Although our patient does not have hypercalcemia, renal dysfunction, or bone lesions to suggest multiple myeloma, amyloidosis is an important cause to consider, and Congo red staining should be considered on biopsy of affected tissue. Similarly, serum free light chains and SPEP/UPEP followed by immunofixation should be obtained. Our patient's SPEP/UPEP with immunofixation showed no monoclonal protein, but serum free light chains were not assessed.

Giant cell arteritis and polymyalgia rheumatica. Giant cell arteritis (GCA) is a large vessel vasculitis that can affect patients over the age of 50 years. It is widely considered to be part of the same disease spectrum as polymyalgia rheumatica (PMR).[36] Several features of this case support the diagnosis: months of constitutional symptoms with markedly elevated

markers of inflammation, jaw claudication, transient vision loss consistent with amaurosis fugax (a hallmark feature caused by ischemia to the retina or optic nerve), and dramatic improvement within hours of steroid administration. Although the severe bilateral thigh pain is consistent with a PMR distribution, the finer details, including the lack of morning stiffness and improvement with rest rather than activity throughout the day, are less consistent. Sinusitis is also not a typical manifestation. Critically, the bilateral cortical hypoenhancing renal lesions remain unexplained. Although GCA can affect the abdominal aorta and its branches, renal involvement typically manifests as arterial stenosis rather than parenchymal lesions, and the CTA showed no aortitis or renal artery stenosis. The negative temporal artery ultrasound finding does not exclude GCA given imperfect sensitivity, particularly after glucocorticoid initiation.[37] Even if it were positive, a TAB should still be pursued given the patient's atypical multiorgan presentation.[38]

ANCA-associated vasculitis. The positive c-ANCA result with elevated PR3 antibody levels raises strong suspicion for ANCA-associated vasculitis, particularly granulomatosis with polyangiitis (GPA). Constitutional symptoms and myalgias are common. Our patient's history of chronic sinusitis could represent upper respiratory tract involvement, a hallmark of GPA. Critically, ANCA-associated vasculitis can rarely cause temporal arteritis through inflammation of the vasa vasorum.[39] Pauci-immune glomerulonephritis is the typical renal manifestation of GPA; however, renal pseudotumors consisting of granulomas, neutrophilic infiltrates, and persistent ischemic changes may appear as renal lesions.[40,41] The absence of asthma and a normal eosinophil count decrease the clinical suspicion for eosinophilic GPA.[42] Microscopic polyangiitis is unlikely without pulmonary involvement, glomerulonephritis, and typical MPO/perinuclear ANCA association.[43] GPA has significant heterogeneity, and an atypical presentation potentially unifies the findings with sinusitis, temporal arteritis, renal lesions, constitutional symptoms, myalgias, jaw claudication, amaurosis fugax, and bilateral renal lesions. Further tissue confirmation should be pursued.

CLINICAL COURSE

The patient presented with new-onset acute severe bilateral thigh pain, bilateral jaw claudication, and a transient episode of left-sided painless monocular vision loss, along with chronic fatigue, constitutional symptoms, generalized weakness, and sinusitis. Abnormal laboratory findings included normocytic anemia, thrombocytosis, elevated markers of inflammation, and positive c-ANCA and PR3 antibodies. The temporal artery ultrasound obtained was normal. CTA of the chest, abdomen, and pelvis identified persistent bilateral renal cortical asymmetric hypoenhancing lesions. Following the CTA, a TAB and renal biopsy were pursued. The TAB was completed first, showing transmural inflammation (TMI) most pronounced in the adventitia with focal transmural extension consisting of a mixture of T cells, B cells, macrophages, and notable eosinophils, consistent with active arteritis (Figure 2). There was no evidence of fibrinoid necrosis, granulomatous inflammation, or giant cell infiltrate. A subsequent renal biopsy revealed dense tubulointerstitial inflammation and fibrosis with a focally storiform pattern accompanied by a lymphoplasmacytic infiltrate with mixed neutrophils and eosinophils, alternating with areas of normal architecture. Immunohistochemical staining showed numerous IgG-positive polyclonal plasma cells, with only a small minority (<10 per high-power field) positive for IgG4. There was no significant direct immunofluorescence

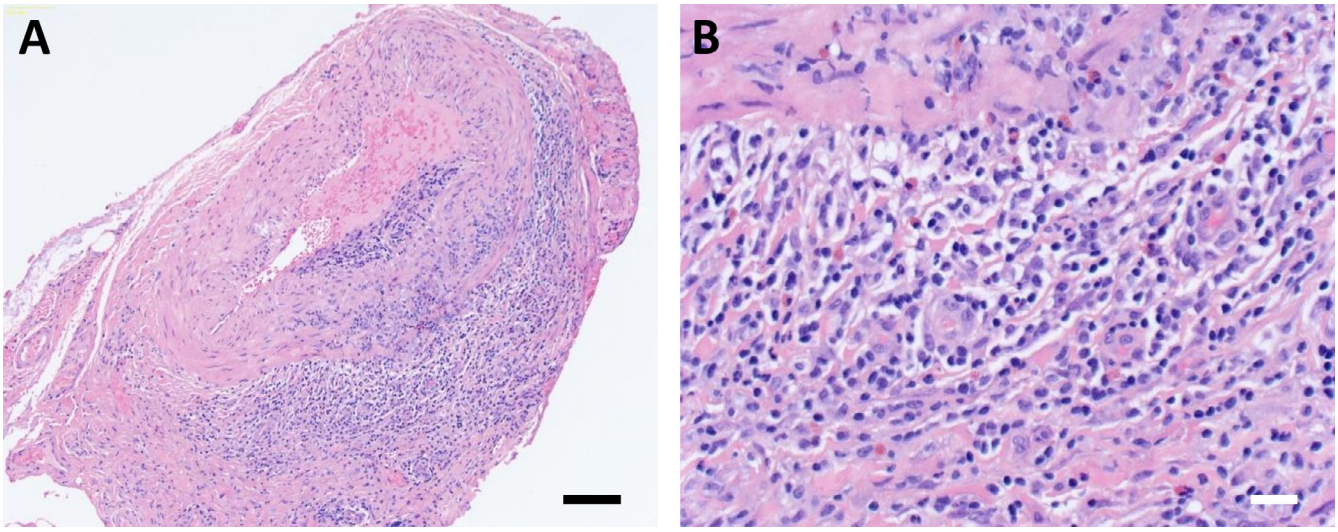

Figure 2. Temporal artery biopsy shows transmural inflammation composed of a mixture of lymphocytes, macrophages, and eosinophils. (A) Scale = 100 μm. (B) Scale = 20 μm.

staining of the glomeruli, crescent formation, or capillary loop basement membrane abnormalities. Collectively, in the context of positive ANCA titers, the patchy but severe tubulointerstitial fibroinflammatory lesions were highly suggestive of renal infiltrative ANCA-associated vasculitis (Figure 3). Results of bacterial, fungal, and mycobacterial cultures from the renal biopsy were negative.

On the day of admission, the patient was empirically started on oral prednisone 60 mg daily for possible GCA with presumed PMR. After she was diagnosed with GPA rather than GCA, she was transitioned to the Plasma Exchange and Glucocorticoids in Severe ANCA-Associated Vasculitis (PEXIVAS trial) reduced prednisone taper and was started on rituximab for induction and maintenance therapy. Her bilateral thigh pain, jaw claudication, generalized weakness, and constitutional symptoms have resolved. Her sinusitis has also significantly improved, although she does continue to endorse milder, ongoing symptoms. The vision loss has not recurred. The chronic, intermittent headaches remain unchanged. A repeat CTA of the abdomen and pelvis is planned to reassess the bilateral renal lesions.

CASE SUMMARY

The patient, a 63-year-old woman, presented with new-onset acute severe bilateral thigh pain, bilateral jaw claudication, and a transient episode of left-sided painless monocular vision loss, along with chronic fatigue, constitutional symptoms, generalized weakness, and sinusitis. Notable laboratory findings included elevated markers of inflammation, positive c-ANCA and PR3 antibodies, normal Ig levels, normal urine studies, and a

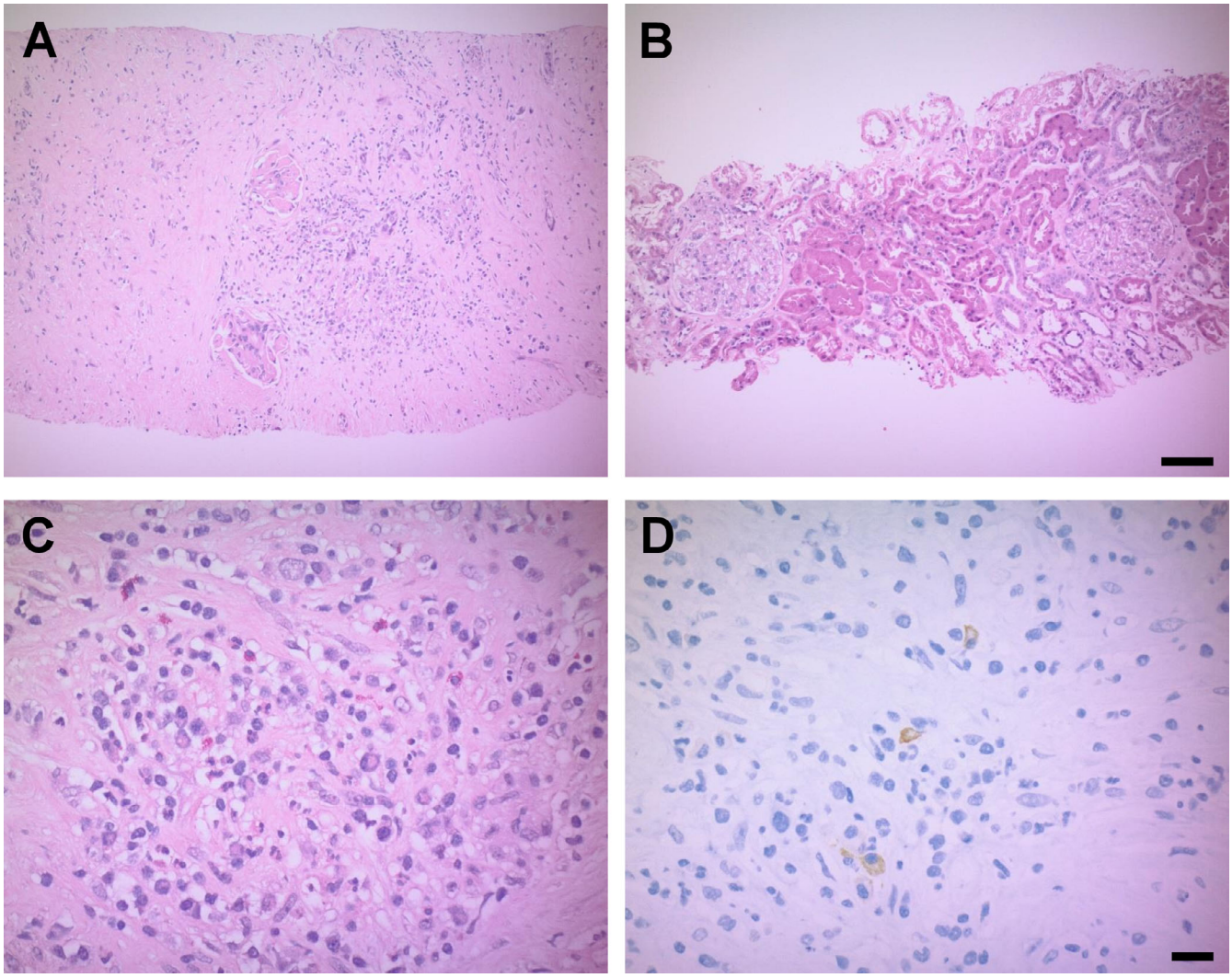

Figure 3. Needle core biopsy of renal lesion. (A) Severe interstitial fibrosis and patchy, focally dense inflammation. Two partially sclerotic glomeruli are present (medium magnification; hematoxylin and eosin stain). (B) Another area showing normal architecture containing two unremarkable glomeruli (medium magnification; hematoxylin and eosin stain). Scale = 100 μm in panels A and B. (C) Mixed inflammatory infiltrate containing numerous plasma cells as well as eosinophils and neutrophils (high magnification; hematoxylin and eosin stain). (D) A few (<10 per high-power field) scattered immunoreactive plasma cells (high magnification; immunohistochemical stain for IgG4). Scale = 20 μm in panels C and D. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80019/abstract.

negative infectious workup result. CTA of the chest, abdomen, and pelvis identified persistent bilateral renal cortical asymmetric hypoenhancing lesions. Subsequent TAB revealed temporal arteritis, and renal biopsy showed patchy tubulointerstitial fibroinflammatory lesions highly suggestive of renal infiltrative ANCA-associated vasculitis. Overall, the case is consistent with GPA.

DISCUSSION

GPA is a systemic, necrotizing, and predominantly small and medium vessel vasculitis. The classic clinical phenotype involves three anatomic sites, the upper respiratory tract, lungs, and kidneys, with the latter manifesting as a pauci-immune necrotizing glomerulonephritis.[44] However, GPA is known to have a heterogeneous clinical profile, presenting with constitutional symptoms, and can also involve the central and peripheral nervous system and musculoskeletal, mucocutaneous, ocular, and, rarely, cardiac and gastrointestinal systems.[45] In this case, positive c-ANCA with markedly elevated PR3 antibody levels (titer 1:160, PR3 antibody level >8.0U/mL), along with confirmatory TAB and renal biopsy findings, established the diagnosis of GPA presenting with a rare phenotype characterized by the combination of temporal arteritis and bilateral renal infiltrative lesions. Other associated clinical manifestations included chronic sinusitis, constitutional symptoms, generalized weakness, myalgias, and jaw claudication and amaurosis fugax secondary to temporal arteritis. Temporal arteritis in GPA has a reported prevalence of 0.8% to 3% of cases but is potentially underreported.[46–48] Similarly, infiltrative renal lesions in GPA are exceedingly rare, with 43 reported cases.[49,50] Upon review of the published literature, this combination has not been previously described in GPA.

In the context of pathology, it is crucial to highlight that using TAB findings to differentiate between causes of primary vasculitis has limitations. TMI, the hallmark of a positive TAB finding suggestive of active arteritis, can be seen in multiple different vasculitides in addition to GCA.[46] Giant cells and granulomatous inflammation can increase diagnostic concern for GCA, but these findings are also not pathognomonic and can be seen in other processes, including GPA. It has been suggested that certain histologic features, especially fibrinoid necrosis and mixed inflammatory infiltrates with significant neutrophils and eosinophils or small vessel vasculitis with inflammation restricted to adventitial or periadventitial small vessels, are less typical for GCA, and a different vasculitis should be considered.[46] In our case, the TMI indeed consisted of a mixed inflammatory infiltrate with notable eosinophils and was most pronounced in the adventitia with focal transmural extension. No fibrinoid necrosis was identified.

The renal biopsy findings in our case were highly suggestive of renal infiltrative ANCA-associated vasculitis. The patchy distribution and extensive plasma cell–rich inflammation with a pattern of interlacing fascicles (“storiform fibrosis”) are characteristic. These findings overlap with IgG4-RD, but there are several features that help distinguish between the two disorders. The presence of interstitial neutrophils and necrosis/karyorrhexis favors a diagnosis of renal infiltrative ANCA-associated vasculitis (Figure 3). IgG4-positive plasma cells can be numerous in both disorders, but they are not always present or prominent in renal infiltrative ANCA-associated vasculitis[51]; their paucity in the current biopsy sample effectively rules out IgG4-RD.[52] Pauci-immune necrotizing glomerulonephritis has been reported in over 50% of patients with renal infiltrative ANCA-associated vasculitis, but its presence is not required for the diagnosis. Granulomatous inflammation and cellular or fibrocellular crescents, which can also be seen, were not identified.[51]

This case emphasizes the diversity with which GPA can present, resulting in significant diagnostic challenges because symptoms are often nonspecific and require deliberate consideration of GPA to prompt the necessary workup. In our case, the patient was initially suspected to have GCA and PMR, but recurrent sinusitis and renal infiltration were inconsistent with GCA, which prompted additional testing, including a thorough vasculitis workup. The patient did not have any localizing abdominal symptoms, which may have deterred comprehensive imaging; however, in pursuit of identifying possible areas of systemic vascular involvement, CTA of the chest, abdomen, and pelvis eventually revealed the bilateral renal lesions. It is also important to mention that without these findings, very likely only a TAB would have been identified as a diagnostic target.

Ultimately, this case highlights the diagnostic mimicry between different types of primary systemic vasculitis, which eventually led to a GPA diagnosis with a rare phenotypic presentation. It emphasizes the importance of comprehensive imaging when considering systemic vasculitis, as patients may not have overt symptoms in areas of anatomic involvement. It also highlights GPA as a cause of renal infiltrative lesions, especially with other supporting features. Finally, the need for histologic support for diagnosis should always be carefully considered, as there can be significant overlapping features across disease domains.

FINAL DIAGNOSIS

Granulomatosis with polyangiitis presenting as bilateral renal infiltrative lesions and temporal arteritis.

ACKNOWLEDGMENTS

There was minimal use of the large language model (LLM) Claude, limited only to checking spelling and grammar in the creation of the manuscript. All other elements, including research of the literature, writing, and content editing of the manuscript, were completed without any use of an LLM. This clinicopathologic conference case article and submission did not require approval from an institutional review board or ethics committee.

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 Shen 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. Hu W, Wang X, Su G. Infective endocarditis complicated by embolic events: pathogenesis and predictors. Clin Cardiol 2021;44(3):307–315. doi:https://doi.org/10.1002/clc.23554
2. Álvarez-Zaballos S, Vázquez-Alen P, Muñoz P, et al. Prevalence and prognostic impact of stroke in a national cohort of infective endocarditis. Int J Stroke 2024;19(9):989–998. doi:https://doi.org/10.1177/17474930241255560
3. Chambers HF, Bayer AS. Native-valve infective endocarditis. N Engl J Med 2020;383(6):567–576. doi:https://doi.org/10.1056/NEJMcp2000400
4. Zumla A, Raviglione M, Hafner R, et al. Tuberculosis. N Engl J Med 2013;368(8):745–755. doi:https://doi.org/10.1056/NEJMra1200894
5. Figueiredo AA, Lucon AM, Srougi M. Urogenital tuberculosis. Microbiol Spectr 2017;5(1):5.1.01. doi:https://doi.org/10.1128/microbiolspec.TNMI7-0015-2016
6. Dhiman R, Lakra S, Panda PK, et al. Neuro-ophthalmic manifestations of tuberculosis. Eye (Lond) 2022;36(1):15–28. doi:https://doi.org/10.1038/s41433-021-01619-6
7. Whitworth HS, Badhan A, Boakye AA, et al; Interferon-γ Release Assays for Diagnostic Evaluation of Active Tuberculosis study group. Clinical utility of existing and second-generation interferon-γ release assays for diagnostic evaluation of tuberculosis: an observational cohort study. Lancet Infect Dis 2019;19(2):193–202. doi:https://doi.org/10.1016/S1473-3099(18)30613-3
8. Furtado JM, Simões M, Vasconcelos-Santos D, et al. Ocular syphilis. Surv Ophthalmol 2022;67(2):440–462. doi:https://doi.org/10.1016/j.survophthal.2021.06.003
9. Cordato DJ, Djekic S, Taneja SR, et al. Prevalence of positive syphilis serology and meningovascular neurosyphilis in patients admitted with stroke and TIA from a culturally diverse population (2005-09). J Clin Neurosci 2013;20(7):943–947. doi:https://doi.org/10.1016/j.jocn.2012.08.011
10. Tiecco G, Degli Antoni M, Storti S, et al. A 2021 update on syphilis: taking stock from pathogenesis to vaccines. Pathogens 2021;10(11):1364. doi:https://doi.org/10.3390/pathogens10111364
11. Katayama K, Ishizuka K, Tawara J, et al. Temporal arteritis caused by tertiary syphilis. Intern Med 2023;62(7):1095–1097. doi:https://doi.org/10.2169/internalmedicine.9779-22
12. Spaltenstein M, Humbert F, Vu DL, et al. A case report of CT-diagnosed renal infarct secondary to syphilitic aortitis. BMC Infect Dis 2017;17(1):520. doi:https://doi.org/10.1186/s12879-017-2624-1
13. Patterson TF, Thompson GR III, Denning DW, et al. Practice guidelines for the diagnosis and management of aspergillosis: 2016 update by the Infectious Diseases Society of America. Clin Infect Dis 2016;63(4):e1–e60. doi:https://doi.org/10.1093/cid/ciw326
14. Cornely OA, Alastruey-Izquierdo A, Arenz D, et al; Mucormycosis ECMM MSG Global Guideline Writing Group. Global guideline for the diagnosis and management of mucormycosis: an initiative of the European Confederation of Medical Mycology in cooperation with the Mycoses Study Group Education and Research Consortium. Lancet Infect Dis 2019;19(12):e405–e421. doi:https://doi.org/10.1016/S1473-3099(19)30312-3
15. Zhu Q, Zhu W, Wu J, et al. Imaging features of primary renal lymphoma. Acta Radiol 2018;59(1):114–120. doi:https://doi.org/10.1177/0284185117706202
16. Attieh RM, Begum F, Chitty D, et al. Kidney and urinary tract involvement in chronic myelomonocytic leukemia. Kidney Med 2024;6(2):100769. doi:https://doi.org/10.1016/j.xkme.2023.100769
17. Shahin OA, Ravandi F. Myeloid sarcoma. Curr Opin Hematol 2020;27(2):88–94. doi:https://doi.org/10.1097/MOH.0000000000000571
18. Morel A, Meuleman MS, Moktefi A, et al. Renal diseases associated with hematologic malignancies and thymoma in the absence of renal monoclonal immunoglobulin deposits. Diagnostics (Basel) 2021;11(4):710. doi:https://doi.org/10.3390/diagnostics11040710
19. Büttner-Herold M, Sticht C, Wiech T, et al. Renal disease associated with myeloproliferative neoplasms and myelodysplastic syndrome/myeloproliferative neoplasms. Histopathology 2021;78(5):738–748. doi:https://doi.org/10.1111/his.14282
20. Hobbs M, Fonder A, Hwa YL. Waldenström macroglobulinemia: clinical presentation, diagnosis, and management. J Adv Pract Oncol 2020;11(4):381–389. doi:https://doi.org/10.6004/jadpro.2020.11.4.5
21. Gertz MA. Acute hyperviscosity: syndromes and management. Blood 2018;132(13):1379–1385. doi:https://doi.org/10.1182/blood-2018-06-846816
22. Minnema MC, Kimby E, D'Sa S, et al. Guideline for the diagnosis, treatment and response criteria for Bing-Neel syndrome. Haematologica 2017;102(1):43–51. doi:https://doi.org/10.3324/haematol.2016.147728
23. Hepschke JL, Ramalingam S, Pohl U, et al. A case of Bing-Neel syndrome presenting like giant cell arteritis. J Neuroophthalmol 2023;43(4):575–579. doi:https://doi.org/10.1097/WNO.0000000000001980
24. Drent M, Crouser ED, Grunewald J. Challenges of sarcoidosis and its management. N Engl J Med 2021;385(11):1018–1032. doi:https://doi.org/10.1056/NEJMra2101555
25. Bergner R, Löffler C. Renal sarcoidosis: approach to diagnosis and management. Curr Opin Pulm Med 2018;24(5):513–520. doi:https://doi.org/10.1097/MCP.0000000000000504
26. Katz G, Stone JH. Clinical perspectives on IgG4-related disease and its classification. Annu Rev Med 2022;73(1):545–562. doi:https://doi.org/10.1146/annurev-med-050219-034449
27. Maleszewski JJ, Tazelaar HD, Horcher HM, et al. IgG4-related disease of the aortic valve: a report of two cases and review of the literature. Cardiovasc Pathol 2015;24(1):56–59. doi:https://doi.org/10.1016/j.carpath.2014.08.001
28. Kondo A, Ikeguchi R, Shirai Y, et al. Association of IgG4-related arteritis with recurrent stroke. J Stroke Cerebrovasc Dis 2020;29(2):104514. doi:https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104514
29. Stone JH, Zen Y, Deshpande V. IgG4-related disease. N Engl J Med 2012;366(6):539–551. doi:https://doi.org/10.1056/NEJMra1104650
30. Ghinai RA, Mahmood S, Mukonoweshuro P, et al. Diagnosing light chain amyloidosis on temporal artery biopsies for suspected giant cell arteritis. J Neuroophthalmol 2017;37(1):34–39. doi:https://doi.org/10.1097/WNO.0000000000000447
31. Muchtar E, Derudas D, Mauermann M, et al. Systemic immunoglobulin light chain amyloidosis-associated myopathy: presentation, diagnostic pitfalls, and outcome. Mayo Clin Proc 2016;91(10):1354–1361. doi:https://doi.org/10.1016/j.mayocp.2016.06.027

32. Naidoo YS, Gupta R, Sacks R. A retrospective case review of isolated sinonasal amyloidosis. J Laryngol Rhinol Otol 2012;126(6):633–637. doi:https://doi.org/10.1017/S0022215112000503
33. Arias M, Abreu JA, Iglesias A, et al. Primary amyloidosis presenting as renal infarction. Eur Radiol 1996;6(3):346–348. doi:https://doi.org/10.1007/BF00180607
34. Thomas J, Ludwig DR, Ballard DH, et al. Spilling the beans: an inside scoop on the imaging of renal parenchymal disease. Abdom Radiol (NY) 2022;47(7):2420–2441. doi:https://doi.org/10.1007/s00261-022-03540-2
35. Rajkumar SV, Dispenzieri A, Kyle RA. Monoclonal gammopathy of undetermined significance, Waldenström macroglobulinemia, AL amyloidosis, and related plasma cell disorders: diagnosis and treatment. Mayo Clin Proc 2006;81(5):693–703. doi:https://doi.org/10.4065/81.5.693
36. Tomelleri A, van der Geest KSM, Khurshid MA, et al. Disease stratification in GCA and PMR: state of the art and future perspectives. Nat Rev Rheumatol 2023;19(7):446–459. doi:https://doi.org/10.1038/s41584-023-00976-8
37. Buttgereit F, Dejaco C, Matteson EL, et al. Polymyalgia rheumatica and giant cell arteritis: a systematic review. JAMA 2016;315(22):2442–2458. doi:https://doi.org/10.1001/jama.2016.5444
38. Schmidt WA. The ultrasound halo sign of temporal arteries: is it always giant cell arteritis? Rheumatology (Oxford) 2019;58(11):1898–1899. doi:https://doi.org/10.1093/rheumatology/kez355
39. Motam A, Khalid A, Elhadi M. Granulomatosis with polyangiitis mimicking giant cell arteritis. Cureus 2025;17(2):e78502. doi:https://doi.org/10.7759/cureus.78502
40. Higashihara M, Kawamura T, Miyazaki Y, et al. Granulomatosis with polyangiitis presenting as a solitary renal mass: a case report. BMC Nephrol 2023;24(1):223. doi:https://doi.org/10.1186/s12882-023-03276-1
41. Fu Z, Liu M, Li Z, et al. Occult renal granulomatous inflammatory lesions in granulomatosis with polyangiitis detected by 18F-FDG PET/CT. Clin Nucl Med 2017;42(9):707–708. doi:https://doi.org/10.1097/RLU.0000000000001746
42. Solans-Laqué R, Rúa-Figueroa I, Blanco Aparicio M, et al. Red flags for clinical suspicion of eosinophilic granulomatosis with polyangiitis (EGPA). Eur J Intern Med 2024;128:45–52. doi:https://doi.org/10.1016/j.ejim.2024.06.008
43. Kronbichler A, Bajema IM, Bruchfeld A, et al. Diagnosis and management of ANCA-associated vasculitis. Lancet 2024;403(10427):683–698. doi:https://doi.org/10.1016/S0140-6736(23)01736-1
44. West SG, Kolfenbach J. Rheumatology Secrets. 4th ed. Elsevier; 2019.
45. Comarmond C, Cacoub P. Granulomatosis with polyangiitis (Wegener): clinical aspects and treatment. Autoimmun Rev 2014;13(11):1121–1125. doi:https://doi.org/10.1016/j.autrev.2014.08.017
46. Cavazza A, Muratore F, Boiardi L, et al. Inflamed temporal artery: histologic findings in 354 biopsies, with clinical correlations. Am J Surg Pathol 2014;38(10):1360–1370. doi:https://doi.org/10.1097/PAS.0000000000000244
47. Nishino H, DeRemee RA, Rubino FA, et al. Wegener's granulomatosis associated with vasculitis of the temporal artery: report of five cases. Mayo Clin Proc 1993;68(2):115–121. doi:https://doi.org/10.1016/S0025-6196(12)60157-8
48. Coattrenec Y, Muller YD, Spoerl D, et al. Prevalence of large vessel vasculitis in ANCA-associated vasculitis: a retrospective cohort study. Rheumatol Int 2021;41(12):2147–2156. doi:https://doi.org/10.1007/s00296-021-04993-2
49. Ahmed N, Rashid S, Warrington KJ, et al. Systemic granulomatosis with polyangiitis presenting with bilateral renal masses. J Rheumatol 2025;52(7):731. doi:https://doi.org/10.3899/jrheum.2024-1177
50. Jordà-Sánchez R, Salvador R, Prieto-González S. A man with fever, retinal artery occlusion, atrioventricular block and bilateral renal masses. Eur J Intern Med 2025;141:106439. doi:https://doi.org/10.1016/j.ejim.2025.106439
51. Gilani SI, Alexander MP, Nasr SH, et al. Clinicopathologic findings in mass forming ANCA-associated vasculitis. Kidney Int Rep 2022;7(12):2709–2713. doi:https://doi.org/10.1016/j.ekir.2022.09.019
52. Gilani SI, Buglioni A, Cornell LD. IgG4-related kidney disease: clinicopathologic features, differential diagnosis, and mimics. Semin Diagn Pathol 2024;41(2):88–94. doi:https://doi.org/10.1053/j.semdp.2023.12.001

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1098–1100
DOI 10.1002/acr.70026
© 2025 American College of Rheumatology

EDITORIAL

Common Data Elements in Autoimmune Disease Research

Victoria K. Shanmugam, MBBS, FRCP, FACR, FACP, CCD,[1] Carmen Ufret-Vincenty, PhD,[1] Xinrui Li, PhD,[1] Anne Deslattes-Mays, PhD,[2] Richard H. Scheuermann, PhD,[2] Belinda Seto, PhD,[3] and Susan Gregurick, PhD[3]

Autoimmune diseases encompass more than 140 diseases and conditions in which the immune system breaks self-tolerance and mistakenly attacks healthy organs and tissues. Although many autoimmune diseases are individually rare, collectively autoimmune diseases are some of the most common chronic diseases in the United States. Robust data on prevalence of autoimmune diseases in the United States are limited, but estimates indicate that between 23.5 and 50.0 million people in the United States live with autoimmune diseases.[1,2]

Common Data Elements for research

The development of Common Data Elements (CDEs) is a foundational component of supporting research in all diseases, and the National Institutes of Health (NIH) CDEs Repository, hosted by the National Library of Medicine, provides an online resource for investigators to identify CDEs for their research.[3] CDEs are standardized, precisely defined questions or variables with a specific set of permissible responses for use across multiple studies or datasets to ensure consistent data collection, interoperability, comparability, and reproducibility. They ensure that researchers are consistent when they collect patient characteristics, disease symptoms, and treatment outcomes.

The NIH Data Management and Sharing (DMS) Policy, published in 2023, requires all NIH-funded researchers generating scientific data to submit a robust DMS Plan as part of their application for funding. Through this policy, the NIH aims to foster a culture of data stewardship, ensuring that data are findable, accessible, interoperable, and reusable (FAIR), in alignment with global best practices.[4] CDE development and utilization is a critical step to realizing these FAIR principles and the DMS Policy objectives because it facilitates the mapping of datasets to common data models for research.

Development of consistent CDEs across biomedical research allows data from various studies to be harmonized and increases sample size for statistical analysis. This is essential for aggregation of data from multiple sources to conduct meta-analyses. Facilitating interoperability of datasets is particularly important in rare autoimmune diseases, in which individual researchers might not have access to data from sufficient patients to generate meaningful insights alone. Scientific advances in this area can only be achieved through collaboration across institutions using a distributed data model. With adoption of consistent CDEs, secondary analysis of datasets collected through clinical trials, epidemiologic studies, and genetic research studies becomes more feasible. This broadens the scope, accelerates potential discoveries, and improves return on research investments. By employing CDEs, researchers can ensure that data are collected in a rigorous and reproducible manner across multiple studies. This will ultimately improve data quality, streamline data sharing, and allow more reliable comparisons and analyses across different studies and systems.

Benefits of adoption of CDEs for autoimmune disease research:

- Ensures consistency and comparability of data
- Enriches data discovery and sharing
- Facilitates multicenter collaboration
- Enables meta-analysis and other statistical tools
- Strengthens rigor and reproducibility of research findings
- Ensures interoperability
- Enables increased power by pragmatically permitting data integration

Dr Scheuermann's work was supported by the Intramural Research Program of the National Library of Medicine, NIH, and the Office of Autoimmune Disease Research, NIH (grant ZIA-LM-202402-01). The contributions of the NIH authors are considered Works of the US Government. The findings and conclusions presented in this article are those of the author(s) and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

[1]Office of Autoimmune Disease Research, Office of Research on Women's Health, Office of the Director, NIH, Bethesda, Maryland; [2]National Library of Medicine, NIH, Bethesda, Maryland; [3]Office of Data Science Strategy, NIH, Bethesda, Maryland.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.70026.

Address correspondence via email to at Victoria K. Shanmugam, MBBS, FRCP, FACR, FACP, CCD, at vicki.shanmugam@nih.gov.

Submitted for publication December 1, 2025; accepted in revised form December 3, 2025.

Challenges and solutions to CDE development for autoimmune disease research

In autoimmune disease research, inconsistent data harmonization across studies and biorepositories creates a critical barrier because of limited interoperability of research datasets. Too often, datasets are deposited in general-purpose repositories such as ImmPort,[5] fulfilling the NIH DMS Policy, but without alignment to shared standards, each with its own data dictionary, structure, and logic. This results in a fragmented data landscape, in which data volume cannot compensate for lack of harmonization and inconsistent data standards limit the trust, collaboration, and efficiency that modern science demands. This is particularly problematic in rare autoimmune diseases, in which every dataset counts and the cost of fragmentation leads to slower discovery, duplication of efforts, and reduced return on public investment. Even in more prevalent autoimmune diseases, the fragmented data landscape hinders research into co-occurring autoimmune diseases, as well as studies of commonalities in disease pathogenesis and shared risk factors and biomarkers.

The development of CDEs for autoimmune disease research is essential to overcome these hurdles. Although many datasets exist for autoimmune disease research, coordinated efforts to develop NIH-endorsed CDEs specific to autoimmune disease research have been lagging.

With the establishment of the Office of Autoimmune Disease Research (OADR) in the Office of Research on Women's Health at the NIH, Congress directed the office to coordinate the development of the NIH-Wide Strategic Plan for Autoimmune Disease Research and to identify emerging areas of innovation and research opportunity, coordinate and foster collaborative research across institutes and centers, and provide resources to support planning, collaboration, and innovation.[2] The NIH-Wide Strategic Plan for Autoimmune Disease Research Fiscal Years 2026–2030 emphasizes the importance of developing computational tools, data science standards, and other methods that will support multimodal data-driven approaches for autoimmune disease research and facilitate interoperability of autoimmune disease research datasets.[6]

Congress also recognized the continued need to develop CDEs in disease areas in which they currently do not exist and highlighted complex autoimmune and immune-mediated conditions as a priority. The Office of Data Science and Strategy (ODSS) was encouraged to collaborate with the institutes and centers that oversee research on these conditions to develop CDEs and engage with outside stakeholders in this effort, including professional societies and patient organizations.[7]

Although the benefits of CDEs are clear, the development and implementation of CDEs into existing and new research programs require engagement and consensus building with the many communities engaged in autoimmune disease research. Existing datasets may need to be aligned to a common ontology, which can be a resource-intensive process, but this step is essential to the success of these endeavors. Equally, developing universally accepted standards for complex autoimmune diseases requires input and collaboration across the research ecosystem, including researchers, clinicians, data scientists, patients, caregivers, and regulatory agencies.

NIH initiatives to develop CDEs for autoimmune disease research

Initiatives from the NIH are making significant strides in developing autoimmune disease–focused CDEs. Through intramural cofunding, OADR and ODSS have supported the development of CDEs for myositis research. A crucial aspect of this project involved a series of consensus-building meetings to refine, optimize, and confirm mutual agreement on the selected CDEs. This work has resulted in the development of more than 800 CDEs for submission to the NIH CDEs Repository for endorsement. Engagement of the international myositis community and leading experts in this field, working alongside people living with autoimmune disease and their caregivers, was essential to the success of this endeavor. In addition, the involvement of data science specialists and the use of modern data science approaches, such as Linked Modeling Language,[8] ensured that the developed CDEs are both machine readable and machine actionable. Workflows developed through the myositis CDE development initiative will be made available to investigators in other fields to facilitate CDE development across other autoimmune diseases.

For extramural investigators seeking to build consensus on autoimmune disease CDEs, the NIH Research Conference Grant (R13) offers an opportunity to apply for support for conferences focused on CDE development for autoimmune disease research. CDE development for juvenile idiopathic arthritis is already underway, utilizing the R13 mechanism to support the consensus-building exercises. The NIH ODSS is also supporting an initiative to develop CDEs for inflammatory bowel disease.

To realize the full potential of CDEs, ongoing coordinated efforts are needed across the entire research community, including funding agencies and the journals that publish the research. Education and training in CDE development will help additional disease areas identify and adopt existing CDEs, design new ones, and complete the necessary consensus-building processes to ensure their adoption by researchers and clinicians across the biomedical research continuum.

Conclusions

In conclusion, developing and adopting CDEs is not just a technical necessity; it is a strategic imperative that is vital for advancing autoimmune disease research. By standardizing data

collection and developing or enhancing data models that support collaboration, we will accelerate discovery in autoimmune disease research. This will allow investigators to move beyond simply checking the box that data have been shared and toward a shared goal of collaborating across distributed data systems. Investing in this shared resource has the potential to rapidly advance the science of autoimmunity and ultimately moves everyone toward a unified vision of improving diagnosis, monitoring, treatment, and potentially prevention and cure of autoimmune diseases.

ACKNOWLEDGMENT

The large language model tool Microsoft Copilot was used to assist with spelling and grammar corrections in manuscript preparation.

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 Shanmugam 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. Miller FW. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. Curr Opin Immunol 2023;80:102266. doi:10.1016/j.coi.2022.102266
2. Shanmugam VK, Clayton JA. Introducing the Office of Autoimmune Disease Research. Arthritis Rheumatol 2024;76(8):1214–1217. doi:10.1002/art.42837
3. Villere S. Common Data Elements Repository. Med Ref Serv Q 2024;43(2):182–190. doi:10.1080/02763869.2024.2323896
4. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3(1):160018. doi:10.1038/sdata.2016.18
5. Bhattacharya S, Dunn P, Thomas CG, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data 2018;5(1):180015. doi:10.1038/sdata.2018.15
6. Shanmugam VK, Ufret-Vincenty C, Li X, et al. The inaugural NIH-Wide Strategic Plan for Autoimmune Disease Research (Fiscal Years 2026–2030). Arthritis Rheumatol Published online October 20, 2025. doi:10.1002/art.43411
7. Explanatory Materials for the Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Bill, 2024—H.R. 5894. US House Committee on Appropriations; 2024.
8. Moxon S, Solbrig H, Unni D, et al. The Linked Data Modeling Language (LinkML): A General-Purpose Data Modeling Framework Grounded in Machine-Readable Semantics. Paper presented at: International Conference on Biomedical Ontologies; September 16-18, 2021; Bozen-Bolzano, Italy. Accessed November 24, 2026. https://ceur-ws.org/Vol-3073/paper24.pdf

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1101–1111
DOI 10.1002/acr.80032
© 2026 American College of Rheumatology. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

American College of Rheumatology
Empowering Rheumatology Professionals

Standardized Interoperable Data Collection for Myositis Research: Developing Expert Consensus on Common Data Elements for Myositis Outcome Measures

Didem Saygin, MD,[1] Matthew Diller, PhD,[2] Varsha Surampudi, PhD,[3] Mark Bodkin,[3] Payam Noroozi Farhadi, MD,[4] Christopher A. Mecoli, MD, MS,[5] Audrey Kessel,[3] Rohit Aggarwal, MD, MS,[6] Helene Alexanderson, PhD, RPT,[7] Anthony Amato, MD, PhD,[8] Christie M. Bartels, MD,[9] Olivier Benveniste, MD,[10] Michelle Best,[11] Hector Chinoy, PhD, FRCP, MSc,[12,13] Ingrid de Groot,[14] Brian Feldman, MD, PhD,[15] Adam M. Huber, MD, MSc,[16] Hanna Kim, MD,[17] Susan Kim, MD,[18] Linda Kobert,[19] Valerie Leclair, MD,[20] Manuel Lubinus,[21] Pedro M. Machado, MD, PhD,[22,23,24] Andrew Mammen, MD, PhD,[17] Liza J. McCann, MBBS,[25] Tahseen Mozaffar, MD,[26] Chester Oddis, MD,[6] Julie J. Paik, MD, MS,[5] Angelo Ravelli, MD,[27] Nicolino Ruperto, MD, MPH,[28,29] Jens Schmidt, MD, FEAN,[30] Ellen Werner,[19] Victoria Werth, MD,[31] Adam Schiffenbauer, MD,[4] Richard H. Scheuermann, PhD,[2] and Lisa G. Rider, MD,[4] for the IMACS Myositis CDE Working Group

Objective. Recent progress has been made in developing validated myositis outcome measures. However, critical deficiencies remain for data standardization across myositis registries. Although the National Institutes of Health (NIH) Common Data Elements (CDE) Repository has been developed to facilitate standardized data collection and sharing, few myositis-specific CDEs currently exist. We developed CDEs for myositis outcome measures using novel data science strategies.

Methods. Data dictionaries of myositis registries were examined to understand how outcome measures are currently captured. We used the Linked data Modeling Language, an open-source data modeling framework, to develop computable CDEs. After drafting CDEs for myositis core set measures (CSMs), an international conference was held with an expert myositis panel to reach consensus on the coding of CDEs and prioritize additional measures for CDE creation, using Delphi and modified nominal group techniques. This workflow was repeated for the prioritized measures in the second phase.

Results. A workflow was established for CDE creation. CDEs for 10 myositis CSMs were drafted. After receiving comments to improve their coding, universal agreement among participants was reached for CSM CDEs. The prioritized measures for future CDEs included myositis response and classification criteria, damage measures, physical function measures, and Patient-Reported Outcomes Measurement Information System instruments. CDEs for 18 additional measures were discussed at a second consensus conference. Similarly, high agreement rates were achieved, except for flare criteria. Altogether, 852 new CDEs were created for 27 myositis forms and achieved consensus, readying their deposit in the NIH CDE Repository.

Conclusion. Leveraging multispecialty expertise in myositis and its patient communities and data science expertise of the National Library of Medicine, the first myositis-specific CDEs have been developed to accelerate the ability to conduct interoperable myositis clinical studies and therapeutic trials. The workflow established here should also benefit creation of CDEs and data sharing for other autoimmune diseases.

The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered works of he United States Government. However, the findings and conclusions do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services. The views expressed are those of the authors and not necessarily those of the UK Ntaional Health Service, teh NIHR, or the UK Department of Health.

Supported by the Intramural Research Programs (IRP) of the NIH: the National Institute of Environmental Health Sciences (NIEHS; grant ZIA-ES-101081), the Division of Intramural Research, National Library of Medicine (grant ZIA-LM-202402), and intramural research grant support from the Office of Autoimmune Disease Research/Office of Women's Health and the NIH Office of Data Science Strategies. This project has been funded in part with federal funds from the NIH, Department of Health and Human Services

SIGNIFICANCE & INNOVATIONS

- This project fills a major gap in myositis research by creating the first comprehensive, computable myositis-specific Common Data Elements using novel data science tools.
- This project establishes a rigorous and scalable approach for harmonizing outcome measures that can be leveraged in other autoimmune diseases.

INTRODUCTION

Idiopathic inflammatory myopathies (IIM) are rare systemic autoimmune diseases characterized by chronic inflammation of muscle and are associated with high morbidity.[1] To develop international multidisciplinary consensus on the conduct and reporting of clinical studies in the IIM, the International Myositis Assessment and Clinical Studies Group (IMACS) was established more than 20 years ago.[2] IMACS has achieved many milestones, including development and validation of core set measures (CSMs) of disease activity and damage of muscle, skin, and other target organs in adult and juvenile IIM.[3,4] Additionally, the Paediatric Rheumatology International Trials Organization (PRINTO) developed a set of CSMs for juvenile dermatomyositis (JDM), with a number of measures in common with the IMACS CSMs[5] (Table 1). These CSMs have been widely used in IIM clinical studies and therapeutic trials.[6–8]

Several other milestones toward standardized data collection have been achieved, including development of patient-reported outcome measures (PROMs) in collaboration with the Outcome Measures in Rheumatology (OMERACT) Myositis Working Group[9] and developing consensus standards for conducting IIM therapeutic trials and their design.[10] IIM classification criteria endorsed by the EULAR/American College of Rheumatology (ACR) perform better than previous classification criteria, capturing a broader range of clinical subgroups, and are now preferentially used in IIM clinical studies.[11] ACR/EULAR response criteria for DM, polymyositis (PM), and JDM that are hybrid criteria also provide gradations of response (minimal, moderate, and major clinical improvement).[11] These criteria have been further validated, performing consistently across multiple studies, and are now used as the primary efficacy endpoint for many myositis therapeutic trials. Although significant progress has been made over the last two decades in the development of standardized outcome assessment, there are critical deficiencies for data standardization that hamper the comparability of studies. For example, a common source of discordance between databases is assessment of Patient Global Disease Activity. This assessment is captured using visual analog scales (VAS) at some centers, whereas others use numeric rating scales (NRS) with different levels of precision and units. To work on addressing these challenges, a Data Harmonization Scientific Interest Group has recently been established within IMACS to bring together experts with interest and/or experience in data collection and data sharing across myositis datasets.

The National Institutes of Health (NIH) Common Data Elements (CDE) repository aims to promote research data interoperability, helping researchers share and combine datasets. CDEs are "standardized, precisely defined questions paired with a set of specific allowable responses, used systematically across different sites, studies, or clinical trials to ensure consistent data collection." The NIH CDE Repository (https://cde.nlm.nih.gov/home) contains 1,603 standard CDEs managed by the NIH CDE Working Group and 22,743 CDEs (as of August 9, 2025) compiled into 19 collections. Three of the CDE collections containing 177 CDEs

(contract 75N95021D00012). Social & Scientific Systems was supported under a contract with NIEHS (HHSN2732016000021) to DLH, LLC (formerly Social & Scientific Systems). Drs Saygin and Mecoli's work was supported by NIH IRP Project, Program and Portfolio Support Services contracts (75N98022D00019). Dr H. Kim's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases IRP (AR041215). Dr Machado's work was supported by the NIHR, University College London Hospitals, and Biomedical Research Centre.

[1]Rush University Medical Center, Chicago, Illinois; [2]National Library of Medicine, NIH, Bethesda, Maryland; [3]DLH, LLC, Bethesda, Maryland; [4]National Institute of Environmental Health Sciences, NIH, Bethesda, Maryland; [5]Johns Hopkins University School of Medicine, Baltimore, Maryland; [6]University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; [7]Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden; [8]Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; [9]University of Wisconsin School of Medicine and Public Health, Madison; [10]Sorbonne Université, Hôpital de la Pitié-Salpêtrière, Paris, France; [11]Cure JM Foundation, Leesburg, Virginia; [12]Salford Royal Hospital, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Science Centre, Salford, United Kingdom; [13]The University of Manchester, Manchester, United Kingdom; [14]Spierziekten Nederland, Baarn, the Netherlands; [15]University of Toronto, Toronto, Ontario Canada; [16]IWK Health Centre and Dalhousie University, Nova Scotia, Canada; [17]National Institute of Arthritis Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland; [18]University of California San Francisco; [19]The Myositis Association, Columbia, Maryland; [20]McGill University, Montreal, Quebec, Canada; [21]Myositis Support and Understanding, Lincoln, Delaware; [22]University College London, London, United Kingdom; [23]NIHR University College London Hospitals Biomedical Research Centre, University College London Hospitals NHS Foundation Trust, London, United Kingdom; [24]Northwick Park Hospital, London North West University Healthcare NHS Trust, London, United Kingdom; [25]Alder Hey Children's NHS Foundation Trust, Liverpool, United Kingdom; [26]University of California Irvine School of Medicine; [27]IRCCS Istituto Giannina Gaslini, University of Genoa, Genoa, Italy; [28]Università Milano Bicocca, IRCCS Fondazione San Gerardo dei Tintori, Milano, Italy; [29]Fondazione IRCCS San Gerardo dei Tintori, Fondazione Paediatric Rheumatology International Trials Organization, Monza, Italy; [30]Immanuel University Hospital Ruedersdorf, Brandenburg Medical School, Rudersdorf, Germany; [31]University of Pennsylvania, Philadelphia.

Drs Scheuermann and Rider are co-senior authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80032).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80032.

Address correspondence via email to Lisa G. Rider, MD, at riderl@mail.nih.gov.

Submitted for publication December 8, 2025; accepted in revised form February 20, 2026.

Table 1. Myositis core set activity measures reaching consensus at First Myositis Common Data Elements Consensus Conference*

Myositis measure (reference)	Characteristics	CDE features and key revisions
Common variables	Study ID number, visit number, visit date, assessor, and certification	• Data elements that are reused in every measure • 7 CDEs created
Physician and Patient and Parent Global Disease Activity[a4]	Overall disease activity on 10-cm VAS	• Encode line length and adjustment to 100-mm VAS scale in CDEs • Added 21-point numeric rating scale on form • Ordinal scale made optional on form • 7 CDEs created and 1 reused for PGA and 8 CDEs created for Patient and Parent Global Activity
HAQ and CHAQ[a4]	Assessing activities of daily living in eight domains	• CDEs included scoring of domains, total score and, adjusted scoring to allow for 1–2 missing domains • Clarified use of assistive devices in CDEs • Made Pain VAS optional in CDEs • 49 and 55 CDEs created for HAQ and CHAQ, and 1 reused in each
MMT-8[a4]	Assesses isometric strength in eight proximal, distal, and axial muscle groups on the Kendall scale	• Bundled muscle strength measure, muscle group, dominant side, and side of testing in CDEs • Added questions to specify the dominant side and laterality of test on form • Recommendation to create persistent URL of IMACS website with detailed instructions, and link to it in CDE repository • 9 CDEs created
Myositis Disease Activity Assessment Tool and Extramuscular Global Disease Activity[4]	Assessment of disease activity in seven organ systems, primarily assessing extramuscular activity	• Recommended making separate form for Extramuscular Global Activity by VAS and include the 21-point numeric rating scale • 90 CDEs created and one reused
Muscle Enzymes and Metabolites[a4]	Laboratory results of serum CK, aldolase, AST, ALT, LDH, and creatinine	• Bundled enzyme or metabolite, value, unit of measure, and upper and lower limit of normal values in CDEs • SGOT and SGPT added to naming of CDEs • Results of enzyme tests made optional in CDEs • Use both IU/l and microkat/dl and for creatinine mg/dl and µmol/l to include European units on form • 11 CDEs created
Childhood Myositis Assessment Scale[a4]	Performance-based tool assessing muscle strength, endurance, and physical function in children	• Items and scoring coded in CDEs, including adjusted score to allow for 1–2 missing items • 28 CDEs created
Disease Activity Score[4]	Overall disease activity in JDM, assessing muscle strength and function and skin activity	• Items and scoring coded in CDEs, including adjusted score to allow for 1–2 missing items • 39 CDEs created

* Eleven myositis forms relating to CSMs were voted for consensus at the first Myositis CDE Consensus Conference. Cutaneous Dermatomyositis Disease Area and Severity Index, Derm Life Quality Index, and CHQ-PF50 were also discussed, but not pursued for the CDE repository due to licensing requirements. ALT, alanine transaminase; AST, aspartate transaminase; CDE, Common Data Element; CHAQ, Childhood Health Assessment Questionnaire; CHQ-PF50, Child Health Questionnaire–Parent Form 50; CK, creatine kinase; CSM, core set measure; HAQ, Health Assessment Questionnaire; IMACS, International Myositis Assessment and Clinical Studies Group; JDM, juvenile dermatomyositis; LDH, lactate dehydrogenase; microkat, microkatal; MMT-8, Manual Muscle Testing 8; PGA, Physician Global Activity; SGOT, serum glutamic-oxaloacetic transaminase; SGPT, serum glutamic pyruvic transaminase; VAS, visual analog scale.
[a] Could be relevant to other neuromuscular and/or autoimmune rheumatic diseases.

are "NIH-Endorsed" by the NIH CDE Governance Committee. The criteria for NIH endorsement of CDEs includes (1) clear definition of the variable with a prompt and a response, (2) evidence of reliability and validity, (3) human- and machine-readable format, (4) recommended or designated by a recognized NIH body, and (5) clear licensing and intellectual property status, with preference for open source or for Creative Commons attribution license. Currently, there are no myositis-specific CDEs and few for autoimmune disease in the NIH CDE Repository, although the development of CDEs for autoimmune disease is a high priority.[12] Development and adoption of CDEs representing the CSMs and other validated outcome measures, as well as IIM classification and response criteria,[11] would help address data interoperability in the myositis field, aiding researchers in sharing and combining datasets and further standardizing data collection. This is particularly important for rare and heterogenous diseases such as IIM in which single center cohorts are often small and multicenter collaboration is critical for research. The CDEs developed would provide a standard terminology of concepts, standardized data structures, and independent semantics that would be reusable across physical data models, forms, and datasets and support the findability, accessibility, interoperability, and reusability (FAIR) data principles.[13]

A collection of CDEs for myositis outcome measures and criteria was developed by a working group of myositis researchers,

data scientists, and informaticians. These were presented to a group of myositis experts including rheumatologists, neurologists, dermatologists, and patients and patient advocates from the IMACS community. Their accuracy and use in representing myositis CSMs, other outcome measures, and validated criteria were discussed and vetted in two consensus conferences. These consensus-approved myositis CDEs have now been submitted to the NIH CDE Governance Committee for endorsement and placement in the NIH CDE Repository. The purpose of this report is to inform about the development of these CDEs for myositis as a model for CDE development in other autoimmune diseases.

MATERIALS AND METHODS

CDE development for myositis CSMs. In addition to the IMACS Outcomes Repository, which includes a database of the myositis outcome measures, additional data dictionaries from several myositis databases, including the MYONET registry,[14] JDM Cohort Biomarker Study and Repository,[15] PRINTO, Childhood Arthritis and Rheumatology Research Alliance, INSPIRE-IBM registry, and several university myositis research centers were examined to understand how the measures are coded and captured across different databases (Figure 1). Copyright status and appropriate permissions for all measures were obtained from copyright holders and publishers to deposit the CDEs in the NIH CDE Repository.

The working group developed CDEs for CSMs from 10 IMACS and PRINTO CSMs as well as other myositis outcome measures and EULAR/ACR classification and response criteria. Generally, CDEs are constructed using human-readable labels and descriptions; however, we developed CDEs that are computable, meaning that the CDEs' information and metadata are explicitly captured in a machine-readable format that leverages existing data standards for reuse.[16] The goal of creating computable CDEs is to maximize their findability, accessibility, interoperability, and reusability at scale in accordance with FAIR principles for data sharing, which aligns with the main purpose for creating CDEs.[13,17] Please see Appendix A for project participants.

To develop computable CDEs, we used the Linked data Modeling Language (LinkML), which is an open-source data modeling framework for creating schemas for structured, linked data.[17] We selected this framework because of its compatibility with numerous Semantic Web standards and data formats, the ability to create both machine-readable and machine-actionable schemas, support for reuse and extension of data elements, its ability to support mapping to existing data standards (eg, terminologies and ontologies, including Unified Medical Language System [UMLS] and Observational Medical Outcomes Partnership [OMOP]), and extensive tooling for data validation, conversion, and management. Components of a LinkML schema include classes, slots, and enumerations. A class is a template for organizing databases on shared characteristics, whereas slots and enumerations are templates for

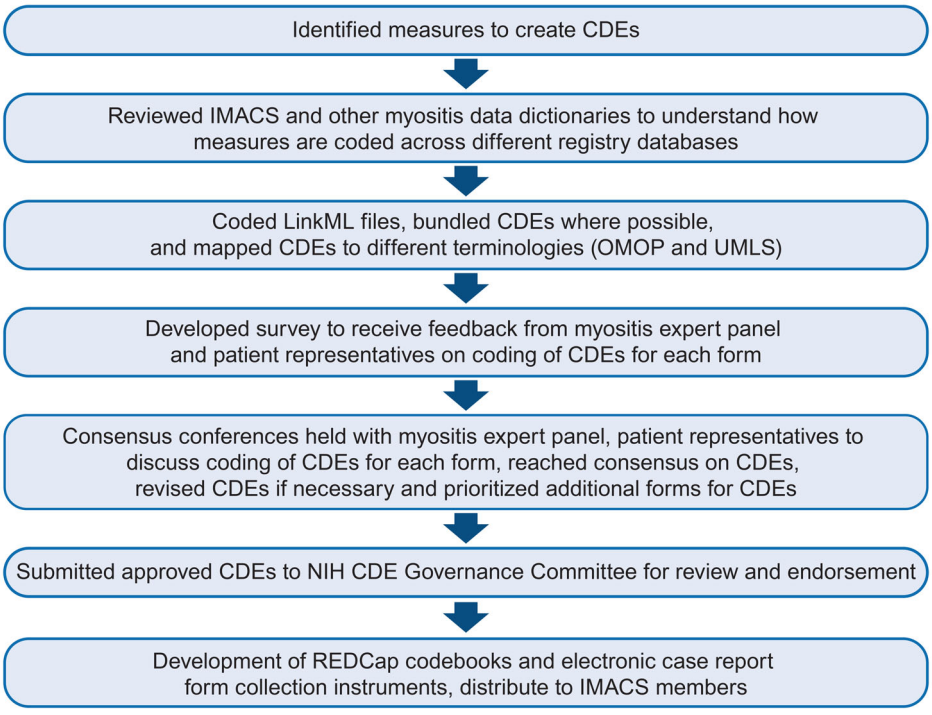

Figure 1. Flow diagram of Myositis Common Data Element Project. CDE, Common Data Element; IMACS, International Myositis Assessment and Clinical Studies Group; LinkML, Linked data Modeling Language; OMOP, Observational Medical Outcomes Partnership; UMLS, Unified Medical Language System.

representing data elements and value lists, respectively. Computable CDEs were developed using a protocol developed in house.

In this myositis CDE project, where possible, CSMs and other fields were bundled based on identified design patterns that are common to each question to reduce the number of CDEs needed for a CSM or other measures to provide a flexible way to combine them for any expanded or short-form versions of the assessment. CDE bundling is a step in the direction of design patterns that supports automation, scalability, reusability, uniform standardized representation, and validation of CDEs. For example, we did not create eight different CDEs for testing the strength of each muscle group for the Manual Muscle Testing 8 (MMT-8) tool; instead, we created one CDE for strength and another for the muscle group assessed. We then reused and bundled this pair of CDEs for each muscle group. This approach reduces redundancy and duplication in the NIH CDE Repository. Examples of myositis forms that use bundled CDEs included MMT-8, Serum Muscle Enzymes and Metabolites, and Timed Up and Go.

Mapping myositis CDEs to standardized terminologies. The Observational Health Data Science and Informatics (OHDSI) web application Athena-OHDSI Vocabularies Repository (Odysseus Data Services, Inc) was used to facilitate mapping from the candidate myositis CDEs to OMOP vocabulary, where they existed (https://athena.ohdsi.org/search-terms/start). Both phrase and exact search parameters were used to identify equivalent OMOP concepts. CDE measures were also mapped to UMLS concepts that were identified as close or exact matches via the UMLS Metathesaurus Browser (https://uts.nlm.nih.gov/uts/umls/home).

Consensus conferences with myositis experts. In the first phase of the project, CDEs were developed for myositis CSMs, and an in-person consensus meeting was held. The goals of the meeting were to discuss the coding of CDEs for myositis CSMs, receive critical feedback from myositis experts, and prioritize additional myositis study forms for creation of new CDEs in the next phase of the project. Key stakeholders were identified and invited to participate, including myositis health care professionals and investigators and patients and patient advocates from IMACS (Supplementary Table 1). This included 13 voting panelists, 11 discussants, and 10 observers in the first consensus meeting. The health professional participants were identified as recognized experts in the myositis field from a variety of specialties and geographic locations with expertise in outcome measures, including those who host, manage, and contribute to local or global myositis databases and those who have led and/or participated in the development of myositis CSMs and/or in clinical trials. The patient and advocate participants were identified through The Myositis Association, Cure JM Foundation, and Myositis Support and Understanding based on their experience in participating in myositis research studies. In the second phase, CDEs were developed for the outcome measures prioritized by the expert panel in the first consensus conference, and a second consensus meeting was held virtually six months later. This study did not involve human subjects and therefore did not require institutional review board approval.

Premeeting surveys. All participants were provided with study forms and instructions for each outcome measure, relevant literature supporting validation, and the CDEs for each of the myositis study forms as CSV or YML files to review before the conference. Participants received an electronic REDCap-based survey to provide feedback on the coding of the CDEs for each myositis measure. The results of the survey were reviewed by the study team before the conference.

Consensus conference methodologies. Consensus sessions involved brief presentations about each myositis measure or criteria set, followed by review of the CDE survey results and voting on the corresponding CDEs for that measure. Two consensus methodologies (Delphi method and modified nominal group technique [mNGT]) were used in tandem to discuss the CDEs for each myositis measure. After each CDE was shared, panelists and discussants were asked to submit any final comments using the REDCap-based survey; survey responses were shared with participants to inform discussions. Any suggested changes to the CDEs agreed to during the discussions were recorded and subsequently incorporated into edited CDE versions following the conference.

This initial step of comments and group discussion was followed by mNGT, which involved formal electronic voting. Consensus was defined as ≥80% agreement among the panelists. If consensus was achieved, then the process proceeded to the next set of CDEs for the next myositis measure without further discussion. If consensus was not achieved (<80% agreement), then a round-robin discussion was held, and each panelist was asked to share their opinion and any comments about the CDEs for a measure, which continued until there were no new comments. The round-robin discussion was followed by a second round of electronic voting to aim to reach consensus.

RESULTS

Consensus on myositis core set disease activity measure CDEs at first conference. CSMs were discussed at the first conference and all achieved consensus, with the panelists endorsing the CDEs as presented and discussed. Altogether, 301 new CDEs were created for 11 forms. The majority of comments related to clarifications of the content of the forms, rather than the coding of the CDEs. Highlights of the discussions are presented here and in Table 1.

A core set of variables. A core set of variables common to every myositis measure was encoded in a single common variables set of CDEs. This included patient identification number, study patient number, assessment number, assessment date,

reporter or assessor, certified by (person checking that the data are complete and within range), and certification date. Based on feedback from the participants, some of these variables were recommended to be made optional.

Physician, Patient, and Parent Global Disease Activity. The Physician, Patient, and Parent Global Disease Activity CDE sets were reused based on overlapping components and similar coding between forms. The CDEs were coded in millimeters to accommodate the integer requirement for electronic capture in the application to a REDCap database. Both vertical marks and line lengths of the VAS were coded. Feedback received from the Delphi survey stated that NRS are currently more frequently used than VAS and have the potential to be more informative (Table 1). One panelist pointed out the infrequent use of the ordinal scale. Two participants recommended adopting only one scale instead of having both ordinal and VAS for the same measure. The group discussion led to a collective decision to make the VAS a required element and the ordinal scale optional and to develop a 21-point NRS for these forms.[18]

Health Assessment Questionnaire and Childhood Health Assessment Questionnaire. Even though the current NIH CDE Repository does not include scoring of forms, this was determined to be an important addition by the study team. Thus, the Health Assessment Questionnaire (HAQ) and Childhood HAQ (CHAQ) CDE sets included domain scores, a total score ranging from 0 to 3, as well as an adjusted score, which allows for one or two missing domains, along with instructions for scoring. The feedback received from the Delphi surveys included a comment to clarify that scoring should be adjusted when a patient uses assistive devices or assistance from another person. Another comment reported a possible difference in the HAQ used in Europe, which generally does not include a pain VAS. Based on group discussions, the pain scale was made optional and the scoring adjusted when assistive devices and aids are used.

MMT. The MMT-8 CDE set was coded both as unilateral using right-sided proximal, distal, and axial muscles with a range of 0 to 80 as originally developed[19] and as a bilateral set with a range of 0 to 150, as is frequently used in therapeutic trials. The CDEs bundled muscle strength measure, muscle group, dominant side, and the side of testing. The coding included a potential score, adjusted score, and a percent score adjusted to a final score with the same denominator (80 or 150) and allowing for missing value(s). The suggestions received from the Delphi survey included adding instructions for performing the testing, specifying the testing side and dominant side, and adding an image or URL of the study form.

Myositis Disease Activity Assessment Tool and Extramuscular Global Disease Activity. The MYOACT CDE VAS scales were coded as 100 mm for all seven organ systems. Twenty-three extramuscular items and three muscle items were categorized as not present, improving, same, worse, and new, as used in the Myositis Disease Activity Assessment Tool (MDAAT).[4] It was suggested that the Extramuscular Global VAS also be completed on an NRS. As the MDAAT, a copyrighted tool, cannot be altered, it was agreed that a separate form would be created for the Extramuscular Global Activity that includes both VAS and NRS scales, which is a CSM for the ACR/EULAR Myositis Response Criteria.[20,21]

Serum muscle enzymes. A data collection form for serum levels of muscle enzymes specified the result and upper limit of normal for each, including creatine kinase (CK), aldolase, aspartate transaminase (AST), alanine transaminase (ALT), and lactate dehydrogenase (LDH), as surrogate laboratory measures of disease activity. Serum creatinine is also included as a measure of muscle atrophy or damage.[3]

The serum muscle enzymes CDE set included coding of each blood laboratory value in IU per liter and microkatal per liter, due to unit differences in the United States and Europe. Upper limit of normal was provided, and values were adjusted to a common upper limit of normal (252 IU/L for CK, 7 IU/L for aldolase, 226 IU/L for LDH, 34 IU/L for AST, and 41 IU/L for ALT) to be able to compare values between centers, regardless of demographic characteristics (age, sex, and race). Minimum and maximum potential values were included. Serum creatinine was encoded in milligram per deciliter and micromole per liter and adjusted to the lower limit of normal of the testing laboratory and a common lower limit of normal of 0.7 mg/dl. In the CDEs for this form, enzyme or metabolite, value, unit of measure, and upper and lower limit of normal values were bundled. Suggestions received from the Delphi survey pointed to the more common use of serum glutamic-oxaloacetic transaminase (SGOT) and serum glutamic pyruvic transaminase (SGPT) terminology, instead of AST and ALT, and the limited availability of aldolase in Europe. Based on this feedback, the enzyme names were updated to include SGOT and SGPT as well as AST and ALT, and the results of each of the laboratory tests were made optional because the availability of the tests varies between institutions. The name of the form was suggested to be changed to Serum Muscle Enzymes and Muscle Metabolites, given that creatinine is not a muscle enzyme. Other CDEs that reached consensus at the first conference included Childhood Myositis Assessment Scale and Disease Activity Scale (Table 1).

At the first consensus conference, two myositis measures were discussed, the Child Health Questionnaire–Parent Form 50 (CHQ-PF50), part of the PRINTO CSMs for JDM, and the Cutaneous DM Disease Area and Severity Index (CDASI), which assesses skin activity and damage. Both tools are licensed with restrictive use, precluding their dissemination through the NIH

CDE Repository. Therefore, these tools were neither coded nor included for voting.

At the end of the first meeting, a survey was conducted to prioritize 44 additional myositis forms as well as the criteria for creation of CDEs at the next phase. The participants were requested to consider the evidence of validation for each form while voting, as required by the NIH CDE Governance Committee (Supplementary Table 2). Fifteen of 25 forms (60%) received at least 30% agreement. Some prioritized forms, including inclusion body myositis (IBM)–related and cancer screening forms were not pursued for CDE creation due to lack of or limited validation studies for these forms. Instead, some forms with <30% agreement, including the Patient-Reported Outcomes Measurement Information System (PROMIS) forms, were prioritized due to the presence of validation studies. Some forms that did not receive high endorsement, including skin global assessment, thigh magnetic resonance imaging, and core elements forms, will be coded in REDCap only, and CDEs would not be created for the NIH CDE Repository (see below).

Consensus on myositis outcome measure and criteria CDEs at the second conference. A second Myositis CDE Consensus Conference was held six months later as a virtual conference with 23 panelists and voting members, almost all of whom participated in the first conference. CDEs for classification and response criteria, damage measures, PROMs, and functional tests were discussed, representing 18 measures. Altogether, 551 new CDEs were created from 17 myositis study forms, which achieved consensus, and 43 CDEs were reused (Table 2).

2017 EULAR/ACR IIM Classification Criteria. CDEs were created for each subclassification category including PM and immune-mediated necrotizing myopathy, DM, IBM, amyopathic DM, JDM, and other juvenile myositis. Scoring was not included. Given the active efforts to update these criteria, some panelists expressed concern for including CDEs for this form; however, the majority felt it was important to include the currently used criteria because it will take a while for new criteria to become available.[11]

IMACS preliminary criteria for disease flare. IMACS CSM CDEs were used as input variables. Absolute and relative percent change between baseline and follow-up values were included. Consensus was not achieved at 50% (10 out of 20); therefore, a round-robin discussion took place, followed by a second vote. Concerns about the form included lack of validation, not being a data-driven criteria but consensus criteria, missing important elements of flare such as pain, fatigue, and skin disease, and not capturing the range and extent of patient symptoms, whereas other panel members thought that the criterion is a good starting point and CDEs should be published. Consensus remained at 52.4% (11 out of 21 participants); therefore, CDEs for disease flare will not be submitted to the NIH CDE Governance Committee.

ACR/EULAR IIM response criteria for DM, PM, and JDM. IMACS CSM CDEs were used as input variables, and all coding elements were initially reused from the disease flare criteria form. Panelists recommended revising the naming of the baseline visit to indicate that it is a prior visit and not necessarily the first patient visit. Because CDEs for the disease flare criteria will not be submitted, the response criteria CDEs were redeveloped, and the CDEs were instead reused from CSM CDEs where possible. New CDEs were developed for absolute change, absolute percentage change, and the distinction between IMACS and PRINTO variables.

Other CDEs that reached consensus at the second conference included Physician and Parent and Patient Global Damage, Myositis Damage Index, PROMIS tools,[9,22] Functional Index-3,[23] 30-second Sit-to-Stand Test, Timed Up and Go test, and Six- and Two-Minute Walk Tests[24] (Table 2).

DISCUSSION

Summary and future directions. Leveraging the broad multispecialty international expertise in myositis and patient communities through the IMACS myositis consortium and the data science expertise of the National Library of Medicine, the first series of myositis-specific CDEs have been developed to accelerate the ability to conduct interoperable myositis clinical studies and therapeutic trials. The data science strategies and methodology used in this study will serve as a paradigm for all autoimmune diseases to develop their disease-specific CDEs. Several of the CDEs, including Physician and Patient and Parent Global Activity and Global Damage, as well as HAQ and CHAQ and PROMIS forms, will have use for other autoimmune diseases, whereas MMT-8, muscle enzymes and metabolites, and functional tests will have broad applicability for other neuromuscular diseases. Additionally, the identification of critical CDEs will help advance database research approaches.

Overall, the panelists achieved consensus on the coding of all but 1 of the 28 myositis study forms, creating 852 CDEs in this project. License requirements create barriers to use of forms, such as CHQ-PF50 and CDASI, and their distribution through the CDE repository. For all the other forms, high rates of agreement could be attributed to the panelists' familiarity with the measures. In fact, most of the panelists had participated in the development of myositis CSMs that served as the foundation for the CDEs, had prior experience with consensus methodologies, and had a track record of participation in other myositis consensus projects, including the development of response criteria and clinical trial design elements. A notable weakness was the absence of CDEs for IBM measures, which largely lacked validation and/or agreement on specific measures, and this should be addressed in the future. Additionally, the CDEs created for

Table 2. Myositis criteria and outcome measures reaching consensus at Second Myositis Common Data Elements Consensus Conference*

Myositis measure (reference)	Characteristics	CDE features and key revisions
EULAR/ACR IIM Classification Criteria[11]	Combination of 16 variables, including age at onset, pattern of weakness, characteristic rashes, laboratory results, and optional muscle biopsy findings. A classification tree defines the myositis subgroup	• CDEs for items and subclassification criteria developed • Clarified that features are present at any time during disease course on form • 47 CDEs created
ACR/EULAR Response Criteria for DM and PM and ACR/EULAR Response Criteria for JDM[20,21]	Hybrid measure derived from IMACS and PRINTO CSMs for DM, PM, and JDM, with a continuous score (Total Improvement Score) and improvement categories	• Reused CDEs from CSMs • Clarified 2 visits used (baseline and follow-up) in the CDEs • 34 CDEs created and 11 reused
Physician and Parent and Patient Global Damage[4]	Overall disease damage on 10-cm VAS, with optional ordinal scale	• Coded similarly to Disease Activity Globals • Added numeric rating scale on form • Made ordinal scale optional on form • Clarified definition of damage for patients and parents on form • 5 CDEs for Physician and 2 CDEs for Patient and Parent Global Damage; 2 CDEs reused in each
Myositis Damage Index[4]	Physician reported; extent and severity of damage of 11 organ systems	• VAS scales and items in 11 organ systems coded in CDEs, along with Severity and Extent of Damage and Extended Damage scores • Clarified damage is captured regardless of etiology and completed every 6 months on form • 159 CDEs created and 5 CDEs reused
PROMIS Short Forms[a]: Fatigue Short-Form 7a, Pain Interference Short-Form 6a, and Physical Function Short-Form 8a[9]	PROMs to assess fatigue, pain interference, and physical function over past 7 days, with *t*-score normalized for general population	• Mapped question IDs to CADSR for cross-referencing in CDE repository • 7 CDEs for PROMIS Fatigue, 6 for Pain, and 8 for Physical function forms created
PROMIS Forms[a]: PROMIS 57 version 2.0, PROMIS 49 Parent Proxy Profile, and PROMIS 49 Pediatric Profile version 2.0[22]	PROMs with *t*-score normalized for general population. Adult form includes assessment of several domains over past 7 days: Physical Function, Pain Interference, Fatigue, Social Role, Anxiety, Depression, and Sleep Disturbance. Pediatric version includes Peer Relations instead of Social Role	• Mapped question IDs to CADSR for cross-referencing in CDE repository • 57 CDEs for PROMIS 57 and 49 CDEs each for PROMIS 49 created
Functional Index-3[4]	Performance-based tool to assess endurance in proximal upper and lower extremities and axial muscle groups in adult patients with DM or PM	• Number of repetitions for each muscle group, percentage of repetitions relative to maximum possible, reasons for limitations, and total score included in CDEs • Clarified to score dominant side, but if there is a functional limitation, then score on nondominant side on form • 54 CDEs created and 2 reused
30-s Sit-to-Stand Test[a24]	Maximum frequency of arising from chair and returning to seated position within 30 s	• Number of full repetitions, arm or device use, and testing completion included in CDEs • Clarified chair height, with thigh and calf at 90-degree angle and feet touching floor on form • 10 CDEs created and 1 reused
Timed Up and Go Test[a]	Time to stand up from chair, walk 3 meters, and sit down	• Time to complete the test, dominant side, status of test completion, and side of gait aide used included in CDEs • Bundled aide type and side of aide in CDEs • 15 CDEs created and 6 CDEs reused

(Continued)

Table 2. (*Cont'd*)

Myositis measure (reference)	Characteristics	CDE features and key revisions
Six- and Two-Minute Walk Tests[a]	Distance walked in two and six minutes	• Walk distance, laps, heart rate, blood pressure, oxygen saturation, aide use, and constraints in performing the test included in CDEs. Walk speed and gait velocity as calculated variables provided • Clarified lap distance and length of track as 30 meters on form • Clarified oximetry to be measured immediately at end of test on form • 64 CDEs created

* Seventeen myositis measures were voted for consensus at the second conference. Additionally, CDEs for the IMACS Preliminary Criteria for Disease Flare were discussed but not pursued for the CDE repository due to <80% agreement.[25] ACR, American College of Rheumatology; CADSR, Cancer Data Standards Registry and Repository; CDE, Common Data Element; CSM, core set measure; DM, dermatomyositis; IIM, idiopathic inflammatory myopathy; IMACS, International Myositis Assessment and Clinical Studies Group; JDM, juvenile dermatomyositis; PM, polymyositis; PRINTO, Paediatric Rheumatology International Trials Organization; PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcomes Measurement Information System; VAS, visual analog scale.
[a] Could be relevant to other neuromuscular and/or autoimmune rheumatic diseases.

myositis study forms are in English, which could significantly limit the global accessibility and adoption of these tools in non–English-speaking countries. Nevertheless, the lessons learned and the methodology developed in this study could serve as a helpful model for development of CDEs in other languages.

Following similar approaches in other autoimmune diseases should expand the capacity for robust generalizable data sharing for these conditions. Further, as new myositis tools become available in the future, the workflow developed here should serve as a helpful guide. As guidance for groups who plan on conducting similar work in other conditions and myositis, we would like to highlight the importance of several elements that made this project successful. Importantly, this project required substantial time commitment from three teams with complementary expertise, including those with data science expertise, myositis subject matter experts and patient representatives, and a core team who organized and coordinated the efforts. Due to differences in collection of the same forms across different databases, we suggest the subject matter expert team have a wide international representation and include individuals with expertise in developing and/or using these forms. Lastly, we strongly recommend future groups take time at project commencement to understand the licensing and copyright requirements of each instrument and review the literature on validation of the instruments before proceeding with CDE development.

As the next steps of this study, the CDEs achieving consensus in this project are in the process of being submitted for endorsement by the NIH CDE Governance Committee and uploaded to the NIH CDE Repository, which is a publicly available website (https://cde.nlm.nih.gov/home). A second method to facilitate the adoption of the myositis CDEs developed in this project will be to develop REDCap data dictionaries and electronic case report form collection instruments for each of the 27 myositis study forms achieving consensus. REDCap is a secure web-based electronic data capture platform designed for building and managing databases for clinical and translational research studies. Each myositis study form created in REDCap will include individual components of the form as well as the built-in automated scoring. For two forms, CHAQ and MDAAT, we were unable to make the REDCap forms available due to licensure restrictions but will only share their data dictionaries. REDCap codebooks of the CDEs will be available on the IMACS website following deposit in the NIH CDE Repository. The link between the NIH CDE Repository and REDCap will facilitate the dissemination of the myositis CDEs and use by myositis researchers, further promoting data compatibility and data sharing. Additionally, our group is working to include the myositis CDEs in the REDCap Shared Library repository for downloading and use by researchers. Using CDEs and REDCap forms for disease activity and damage, PROMs, other key assessment tools, and criteria developed in this study will facilitate development of new myositis registries for investigators and advance the design and conduct of future therapeutic trials for these rare autoimmune diseases.

ACKNOWLEDGMENTS

We would like to thank IMACS Scientific Committee for their careful review of this manuscript. We acknowledge and thank the NIH Office of Data Science Strategies, NIH Office of Autoimmune Disease Research, and the National Library of Medicine for their support of this work, with special thanks to Drs Belinda Seto, Victoria Shanmugam, Carmen Ufret-Vincenty, and Robin Taylor. We thank Drs Olivier Benveniste, Hector Chinoy, Lisa Christopher-Stine, Pedro M. Machado, Christopher Mecoli, Tahseen Mozaffar, Chester Oddis, Nicola Ruperto, Lucy Wedderburn, and Victoria Werth for sharing data dictionaries from their myositis registries to assist in the formulation of CDEs for the myositis study forms.

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 Rider 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. Lundberg IE, Fujimoto M, Vencovsky J, et al. Idiopathic inflammatory myopathies. Nat Rev Dis Primers 2021;7(1):86. doi:10.1038/s41572-021-00321-x
2. Isenberg DA, Chinoy H, Dimachkie MM, et al. The origins, evolution and future of the International Myositis Assessment and Clinical Studies Group (IMACS). Clin Exp Rheumatol 2022;40(2):214–218. 10.55563/clinexprheumatol/yg743i
3. Miller FW, Rider LG, Chung YL, et al; International Myositis Outcome Assessment Collaborative Study Group. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. Rheumatology (Oxford) 2001;40(11):1262–1273. doi:10.1093/rheumatology/40.11.1262
4. Rider LG, Werth VP, Huber AM, et al. Measures of adult and juvenile dermatomyositis, polymyositis, and inclusion body myositis: Physician and Patient/Parent Global Activity, Manual Muscle Testing (MMT), Health Assessment Questionnaire (HAQ)/Childhood Health Assessment Questionnaire (C-HAQ), Childhood Myositis Assessment Scale (CMAS), Myositis Disease Activity Assessment Tool (MDAAT), Disease Activity Score (DAS), Short Form 36 (SF-36), Child Health Questionnaire (CHQ), physician global damage, Myositis Damage Index (MDI), Quantitative Muscle Testing (QMT), Myositis Functional Index-2 (FI-2), Myositis Activities Profile (MAP), Inclusion Body Myositis Functional Rating Scale (IBMFRS), Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI), Cutaneous Assessment Tool (CAT), Dermatomyositis Skin Severity Index (DSSI), Skindex, and Dermatology Life Quality Index (DLQI). Arthritis Care Res (Hoboken) 2011;63 Suppl 11(0 11):S118–S157. doi:10.1002/acr.20532
5. Ruperto N, Ravelli A, Pistorio A, et al. The provisional Paediatric Rheumatology International Trials Organisation/American College of Rheumatology/European League Against Rheumatism Disease activity core set for the evaluation of response to therapy in juvenile dermatomyositis: a prospective validation study. Arthritis Rheum 2008;59(1):4–13. doi:10.1002/art.23248
6. Aggarwal R, Charles-Schoeman C, Schessl J, et al; ProDERM Trial Group. Trial of intravenous immune globulin in dermatomyositis. N Engl J Med 2022;387(14):1264–1278. doi:10.1056/NEJMoa2117912
7. Oddis CV, Reed AM, Aggarwal R, et al; RIM Study Group. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. Arthritis Rheum 2013;65(2):314–324. doi:10.1002/art.37754
8. Ruperto N, Lovell DJ, Quartier P, et al. Abatacept in children with juvenile idiopathic arthritis: a randomised, double-blind, placebo-controlled withdrawal trial. Lancet 2008;372(9636):383–391. doi:10.1016/s0140-6736(08)60998-8
9. Saygin D, DiRenzo D, Raaphorst J, et al; OMERACT Myositis Working Group. Responsiveness and meaningful thresholds of PROMIS pain interference, fatigue, and physical function forms in adults with idiopathic inflammatory myopathies: report from the OMERACT Myositis Working Group. Semin Arthritis Rheum 2024;64:152339. doi:10.1016/j.semarthrit.2023.152339
10. Oddis CV, Rider LG, Reed AM, et al; International Myositis Assessment and Clinical Studies Group. International consensus guidelines for trials of therapies in the idiopathic inflammatory myopathies. Arthritis Rheum 2005;52(9):2607–2615. doi:10.1002/art.21291
11. Lundberg IE, Tjärnlund A, Bottai M, et al; International Myositis Classification Criteria Project Consortium, the Euromyositis Register, and the Juvenile Dermatomyositis Cohort Biomarker Study and Repository (UK and Ireland). 2017 European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Arthritis Rheumatol 2017;69(12):2271–2282. doi:10.1002/art.40320
12. Congressional Record. December 20, 2022 – Issue: Vol. 168, No. 198 — Daily Edition. Senate Section. https://www.congress.gov/congressional-record/volume-168/issue-198/senate-section/article/S8553-2
13. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016;3:160018. doi:10.1038/sdata.2016.18
14. Lilleker JB, Vencovsky J, Wang G, et al. The EuroMyositis registry: an international collaborative tool to facilitate myositis research. Ann Rheum Dis 2018;77(1):30–39. doi:10.1136/annrheumdis-2017-211868
15. Martin N, Krol P, Smith S, et al. A national registry for juvenile dermatomyositis and other paediatric idiopathic inflammatory myopathies: 10 years' experience; the Juvenile Dermatomyositis National (UK and Ireland) Cohort Biomarker Study and Repository for Idiopathic Inflammatory Myopathies. Rheumatology (Oxford) 2011;50(1):137–145. doi:10.1093/rheumatology/keq261
16. Kush RD, Warzel D, Kush MA, et al. FAIR data sharing: the roles of common data elements and harmonization. J Biomed Inform 2020;107:103421. doi:10.1016/j.jbi.2020.103421
17. Moxon SA, Solbrig H, Unni DR, et al. The Linked data Modeling Language (LinkML): a general-purpose data modeling framework grounded in machine-readable semantics. CEUR Workshop Proceedings 2021;3073:148–151.
18. Filocamo G, Davi S, Pistorio A, et al. Evaluation of 21-numbered circle and 10-centimeter horizontal line visual analog scales for physician and parent subjective ratings in juvenile idiopathic arthritis. J Rheumatol 2010;37(7):1534–1541. doi:10.3899/jrheum.091474
19. Rider LG, Koziol D, Giannini EH, et al. Validation of manual muscle testing and a subset of eight muscles for adult and juvenile idiopathic inflammatory myopathies. Arthritis Care Res (Hoboken) 2010;62(4):465–472. doi:10.1002/acr.20035
20. Rider LG, Aggarwal R, Pistorio A, et al. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate, and major clinical response in juvenile dermatomyositis: an International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation collaborative initiative. Arthritis Rheumatol 2017;69(5):911–923. doi:10.1002/art.40060
21. Aggarwal R, Rider LG, Ruperto N, et al; International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate, and major clinical response in adult dermatomyositis and polymyositis: an International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation collaborative initiative. Ann Rheum Dis 2017;76(5):792–801. doi:10.1136/annrheumdis-2017-211400
22. Austenfeld E, Sabbagh S, Liegl M, et al. Validation of PROMIS in adult and pediatric patients with idiopathic inflammatory myopathies [abstract]. Arthritis Rheumatol 2023;75(suppl 9). https://acrabstracts.org/abstract/validation-of-promis-in-adult-and-pediatric-patients-with-idiopathic-inflammatory-myopathies/
23. Ernste FC, Chong C, Crowson CS, et al. Functional Index-3: a valid and reliable functional outcome assessment measure in patients with

dermatomyositis and polymyositis. J Rheumatol 2021;48(1):94–100. doi:10.3899/jrheum.191374

24. Saygin D, Pillai AC, Wakeley AK, et al. Performance of sit-to-stand, timed up-and-go, and six-minute walk tests in home and office settings in patients with idiopathic inflammatory myopathy. Semin Arthritis Rheum 2025;74:152809. doi:10.1016/j.semarthrit.2025.152809
25. Oddis CV, Rider LG, Reed AM, et al. International consensus guidelines for trials of therapies in the idiopathic inflammatory myopathies. Arthritis Rheum 2005;52(9):2607–2615. doi:10.1002/art.21291

APPENDIX A: PROJECT PARTICIPANTS

The participants who contributed to this project are as follows:

The members of the IMACS Myositis CDE Study Group are as follows: Lisa G. Rider (co-principal investigator, National Institute of Environmental Health Sciences, NIH), Richard H. Scheuermann (co-principal investigator, National Library of Medicine, NIH), Didem Saygin, Matthew Diller, Varsha Surampudi, Mark Bodkin, Payam Noroozi Farhadi, Christopher Mecoli, Audrey Kessel, and Adam Schiffenbauer.

The participants in consensus conferences on Myositis Expert Panel are as follows: Rohit Aggarwal, Helene Alexanderson, Christie Bartels, Olivier Benveniste, Hector Chinoy, Brian Feldman, Adam M. Huber, Hanna Kim, Susan Kim, Valerie Leclair, Andrew Mammen, Liza McCann, Tahseen Mozaffar, Chester Oddis, Julie Paik, Angelo Ravelli, Nicola Ruperto, Jens Schmidt, Victoria Werth, and Pedro Machado.

The participants in consensus conference that were patient and patient advocate representatives were as follows: Michelle Best, Ingrid de Groot, Linda Kobert, Manuel Lubinus, and Ellen Werner Shaller.

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1112–1122
DOI 10.1002/acr.80004
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

American College of Rheumatology
Empowering Rheumatology Professionals

Marked Long-Term Improvement in Lung Function in Melanoma Differentiation–Associated Protein 5 Antibody–Positive Dermatomyositis Patients: Experience of a Single-Center Longitudinal Cohort in North America

Jenice X. Cheah, MD, Sangmee S. Bae, MD, MS, Tiffany De Leon, MBA, Yuna Lee, BS, Rong Guo, MS, David Elashoff, PhD, Jennifer Wang, BS, Ani Shahbazian, BS, and Christina Charles-Schoeman, MD, MS

Objective. The objective of this study was to describe the longitudinal disease course and pulmonary outcomes of North American patients with melanoma differentiation–associated protein 5 (MDA5) antibody–associated dermatomyositis (DM).

Methods. Thirty patients with MDA5 antibody–associated DM were identified in a single-center longitudinal cohort of 352 patients with idiopathic inflammatory myopathies. Longitudinal assessments of patient clinical and laboratory disease characteristics, pulmonary function tests (PFT), and high-resolution computed tomography chest scans were conducted.

Results. Eighty percent (n = 24/30) of patients with MDA5 antibody–associated DM had interstitial lung disease (ILD). The overall mortality was low (2/24 at a mean ± SD follow-up of 4.0 ± 0.8 years). At this follow-up, patients were receiving 3.1 ± 1.3 therapies, including 79% receiving intravenous Ig (IVIg), 58% receiving rituximab, 67% receiving mycophenolate, and 63% receiving glucocorticoids. In 18 of 22 surviving patients with ILD who had two-year longitudinal follow-up available at 1.8 ± 0.6 years, improvements of 16% and 17% predicted forced vital capacity (FVC) and diffusing capacity for carbon monoxide (DLco) were noted. In 10 of 18 patients with additional long-term follow-up available (mean ± SD 6.8 ± 3.4 years), improvements of 24% and 20% predicted FVC and DLco were noted. MDA5 antibody and interleukin-15 (IL-15) levels and paraoxonase 1 (PON1) enzyme activity correlated significantly with disease activity at baseline and longitudinally.

Conclusion. In a North American MDA5 antibody–associated DM-ILD cohort treated with aggressive combination immunomodulatory therapy including predominantly mycophenolate, IVIg, and rituximab, disease mortality was low and lung function improved markedly. IL-15, PON1, and MDA5 antibody titers warrant further investigation as disease activity biomarkers in this high-risk population.

INTRODUCTION

Dermatomyositis (DM) is an idiopathic inflammatory myopathy (IIM) characterized by classic rashes and myopathy, which can associate with interstitial lung disease (ILD).[1–3] Patients with DM with antibodies against melanoma differentiation–associated protein 5 (MDA5) have a higher risk of ILD, and[4–7] studies in Asian cohorts have found that not only is ILD more common in this patient population, but patients with MDA5 antibody–positive DM are also more likely to have rapidly progressive ILD (RP-ILD) with a markedly early mortality.[4–8] The association of MDA5 antibody–positive DM with RP-ILD is less well described in North American cohorts, though it remains a serious complication and is a leading cause of death.[1,5,6] More information is needed regarding the disease course and pulmonary outcomes in North American patients with MDA5 antibody–positive DM.

Vasculopathy is strongly implicated in the pathogenesis of MDA5 antibody–positive DM, although mechanisms that drive this vasculopathy are unclear. Paraoxonase 1 (PON1), a high-density lipoprotein (HDL)–associated enzyme, protects the

University of California, Los Angeles.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80004).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80004.

Address correspondence via email to Christina Charles-Schoeman, MD, MS, at ccharles@mednet.ucla.edu.

Submitted for publication May 22, 2025; accepted in revised form December 18, 2025.

SIGNIFICANCE & INNOVATIONS

- Patients with MDA5 autoantibody–positive dermatomyositis (DM) and interstitial lung disease (ILD) may have marked improvement in pulmonary function.
- Aggressive combination immunotherapy is associated with improved pulmonary outcomes in MDA5 autoantibody–positive DM–associated ILD.
- MDA5 autoantibody–positive DM disease activity is associated with autoantibody titers, interleukin-15 levels, and paraoxonase 1 activity.

vascular endothelium from damage due to inflammatory oxidized phospholipids.[9–11] We previously reported abnormal enzyme activity of PON1 in patients with IIM, which was associated with worse disease activity and severe ILD.[10,11]

Presently, we describe the clinical and laboratory characteristics of patients with MDA5 antibody–positive DM sequentially enrolled in a longitudinal observational study at a single North American academic center. We report longitudinal improvement in lung function in the majority of MDA5 antibody–positive patients with ILD, most of whom had severe disease requiring initial hospitalization and were treated with aggressive combination therapy longitudinally. We further describe associations of improvement in disease activity with decreases in MDA5 antibody and circulating interleukin-15 (IL-15) levels, as well as improvement in the activity of PON1.[10] The current data highlight the importance of early recognition and aggressive combination treatment of MDA5 antibody–positive DM-ILD in North American patients who have the potential for marked longitudinal improvement in lung function.

METHODS

Patients. Consecutive newly and previously diagnosed patients with myositis seen at the University of California, Los Angeles (UCLA) Medical Center who met the EULAR/American College of Rheumatology (ACR) criteria for at least "probable" adult IIM[12] were recruited into a longitudinal IIM cohort. The first 30 consecutively enrolled patients with DM who were MDA5 antibody positive were analyzed in the current study. Written informed consent was obtained from all patients under a protocol approved by the Human Research Subject Protection Committee at UCLA (IRB# 10-001833).

Clinical data collection. Demographic information, medication use, DM disease features, and hospitalization history were obtained from patient chart review and questionnaires. Disease activity and damage were assessed using physician global assessment of myositis disease activity and damage scales via a 100-mm visual analog scale (VAS).

Patients provided blood samples at routine clinical visits and completed questionnaires regarding health information. Laboratory assessments of creatine phosphokinase (CPK), aldolase, high-sensitivity C-reactive protein (hsCRP), the Westergren erythrocyte sedimentation rate (ESR), and lipid profiles were performed using standard methods. Myositis-specific antibodies were assessed using standardized laboratory protocols (14 by immunoprecipitation at the Oklahoma Medical Research Foundation and 16 through other standard clinical laboratory tests using enzyme-linked immunosorbent assay [ELISA] or line blot). Additional blood samples were collected in heparinized tubes for plasma isolation and stored at –80°C for additional assays as described in the following sections.

Assessment of pulmonary function tests. Serial pulmonary function tests (PFTs) completed for routine clinical care in MDA5 antibody–positive patients with ILD were retrospectively reviewed. Patients' initial PFTs closest to the date of DM diagnosis were obtained. The subsequent four PFTs and the most recent PFTs before the study's end date (April 14, 2023) were also recorded. PFT data included forced vital capacity (FVC) and the diffusing capacity for carbon monoxide adjusted for hemoglobin (DL_{CO} Hg). Short-term follow-up was defined as the available PFTs closest to the two-year mark from entry into the cohort. Long-term follow-up was the last PFT available at the time of the analysis (April 14, 2023) after the two-year time point.

High-resolution computed tomography chest imaging. Radiology reports of high-resolution computed tomography (HRCT) chest scans closest to patients' initial PFTs and the most recent HRCT chest scans before the study's end date were obtained. Radiology reports were reviewed for the following characteristics: (1) ground glass opacities or consolidations; (2) reticulations; (3) small airway disease, bronchiectasis, or bronchial wall thickening; and (4) fibrosis or fibrotic changes.

Assessment of immunomodulatory therapies. We assessed immunomodulatory therapies received by patients at DM diagnosis and longitudinally at 3, 6, 9, 12, 18, and 24 months following diagnosis and at the most recent visit before the study's end date.

Quantitative MDA5 antibody and ferritin level assessments. MDA5 antibody titers and ferritin levels were measured in patient plasma samples using standard ELISA kits from MBL International and R&D Systems, respectively.

Determination of circulating cytokine and chemokine levels. Luminex multiplexed immunoassays were performed on patient plasma samples for 37 cytokines, chemokines, and growth factors: eotaxin, G-CSF, GM-CSF, IFNα2, IFNγ, IL-1α, IL-1β, IL-1ra, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8,

IL-10, IL-12(p40), IL-12(p70), IL-13, IL-15, IL-17A, IL-17E/IL-25, IL-17F, IL-18, IL-22, IP-10, MCP-1, M-CSF, MIG, MIP-1α, MIP-1β, PDGF-AB/BB, RANTES, TNFα, TNFβ, and VEGF-A.

PON1 activity. The enzymatic activity of PON1 was quantified using three different substrates—paraoxon, dihydrocoumarin, and phenylacetate—for its paraoxonase, lactonase, and arylesterase activities, as described previously.[13] The PON1 Q192R polymorphism was determined in patients as described previously.[14]

Statistical analysis. Statistical analyses were performed using SAS version 9.4 (SAS Institute Inc). All *P* values were two sided, and significance was denoted at <0.05. For Table 1, standard descriptive statistics were reported, with mean (SD) for numerical variables and n (percentage) for categorical variables. Wilcoxon rank sum tests were used for the comparison of skewed continuous variables, *t*-tests were used for normally distributed continuous variables, and Fisher's exact tests were used for categorical variables. For Figure 1, paired *t*-tests were used for the comparison of PFT data between time points. For Figure 2, mixed-effect models, as well as the generalized estimating equations (GEE) method with a Poisson distribution, were used to examine the effects of time, group (≥20% vs <20% improvement in both FVC and DLco Hg), and their interaction on the number of immunotherapies. For Figure 3, McNemar's tests were performed to compare baseline and follow-up CT findings. For Figures 4 and 5, Spearman's correlation coefficients were calculated to examine the association between the variables of interest.

RESULTS

Clinical characteristics of patients with anti-MDA5 antibody–positive DM. Thirty patients with DM who were MDA5 antibody positive were identified in a single-center longitudinal cohort of 352 patients with IIM (total 8.5% with MDA5 antibody–positive DM). Twenty-four of the MDA5 antibody–positive patients had a diagnosis of ILD, and six patients did not have ILD (Table 1). Patients with DM-ILD had an aggressive disease phenotype, with 18 of 24 requiring initial hospitalization. Of the 18 patients with DM-ILD who were hospitalized early in their diagnosis, 15 of 18 had hospitalizations with respiratory complaints such as shortness of breath and hypoxia. Six patients had aggressive pulmonary disease requiring supplemental oxygen, including two patients who were intubated and died. Three surviving patients no longer required supplemental oxygen after initial treatment, and one remains stable on 2 L of supplemental oxygen after more than 10 years. Patients without DM-ILD were significantly younger and more likely to present with inflammatory arthritis (Table 1).

Patients with DM-ILD in the cohort were treated with combination immunotherapy during longitudinal follow-up and continued receiving combination immunotherapy, with a mean ± SD of 3.1 ± 1.3 individual therapies at the last follow-up visit available for this analysis (Table 1). Patients without DM-ILD received fewer numbers of combined therapies, continuing to receive a mean ± SD combination of 2.0 ± 1.3 therapies at the last follow-up visit (DM-ILD vs non-DM-ILD group, $P = 0.06$).

Longitudinal pulmonary outcomes of patients with anti-MDA5 antibody–positive DM-ILD. Two of the 24 patients with DM-ILD died of rapidly progressive disease within three months of cohort entry (mean ± SD disease duration from diagnosis: 5 ± 2 months). Of the 22 remaining patients, 18 had short-term follow-up PFTs available for review at a mean ± SD of 1.8 ± 0.6 years from cohort entry. The other four patients had recently entered the cohort and did not have PFT follow-up data available. Ten of the surviving 22 patients with DM-ILD also had long-term follow-up data available at approximately seven years (mean ± SD 6.8 ± 3.4 years) from cohort entry. None of the patients underwent lung transplantation.

Over short-term follow-up, pulmonary function significantly improved in the overall DM-ILD cohort. Compared to baseline, our short-term follow-up group had significant absolute improvements in percent predicted DLco Hg and FVC (mean ± SD 16.8% ± 18.8% and 15.9% ± 15.5%, respectively) at two years (Figure 1A, Supplemental Table 1). Our long-term follow-up group (approximately seven years) had even greater improvements in their PFT results, with mean ± SD absolute improvements in percent predicted DLco Hg and FVC of 20.0% ± 15.6% and 24.0% ± 20.9%, respectively, from baseline to long-term follow-up (Figure 1B, Supplemental Table 1).

Combination immunotherapy over longitudinal follow-up is associated with improvement in lung function. Most patients with DM-ILD received combination therapy and continued receiving an average of three therapies at the last follow-up visit, which was a mean ± SD of 4.0 ± 3.7 years from diagnosis (Table 1). Medication use was available for assessment longitudinally in 17 of 18 patients in the short-term follow-up cohort (Figure 1A). These patients were grouped into two categories by FVC and DLco improvements: group A, >20% relative improvement in both FVC and DLco Hg at the two-year follow-up; and group B, ≤20% relative improvement in both FVC and DLco Hg at the two-year follow-up visit. Percent change from baseline PFT values was assessed to account for baseline severity of pulmonary disease.

In the first two years after diagnosis, patients in group A (who showed marked pulmonary improvement) received a greater mean number of combined therapies than those in group B (Figure 2). By two years following diagnosis, the mean number of

Table 1. Clinical and laboratory data for patients with MDA5 antibody–positive DM with or without ILD*

	ILD (n = 24)	No ILD (n = 6)	Overall (n = 30)
Demographics			
Age, years	49.1 ± 15.0	29.8 ± 12.4	45.2 ± 16.3
Sex, female	16 (67)	4 (67)	20 (67)
Ethnicity, Hispanic	6 (25)	1 (17)	7 (23)
Race, White	13 (54)	3 (50)	16 (53)
Clinical characteristics[a]			
Required hospitalization for DM	18 (75)	5 (83)	23 (77)
Inflammatory arthritis	15 (63)	6 (100)	21 (70)
Proximal muscle weakness	21 (88)	4 (67)	25 (83)
Rash	24 (100)	6 (100)	30 (100)
Dysphagia	6 (25)	2 (33)	8 (27)
Calcinosis	4 (17)	1 (17)	5 (17)
Cutaneous ulcerations	9 (38)	1 (17)	10 (33)
Fever	5 (21)	1 (17)	6 (20)
Weight loss	8 (33)	2 (33)	10 (33)
Myalgia	8 (33)	0 (0)	8 (27)
Oral ulcers	6 (25)	0 (0)	6 (20)
Heart failure	0 (0)	1 (17)	1 (3)
Pulmonary hypertension	1 (4)	0 (0)	1 (3)
Raynaud phenomenon	5 (21)	2 (33)	7 (23)
Voice hoarseness	9 (38)	3 (50)	12 (40)
Disease duration at last follow-up, y	4.0 ± 3.7	6.7 ± 4.5	4.5 ± 4.0
Laboratory/disease activity characteristics[b]			
Physician global assessment of disease activity (VAS)	51.5 ± 23.3	37.2 ± 13.7	48.6 ± 22.3
CPK, U/L	61.3 ± 35.8	68.5 ± 70.9	62.8 ± 43.5
LDH, U/L	286.4 ± 102.0	252.5 ± 88.3	280.2 ± 98.6
Aldolase, U/L	6.2 ± 2.2	6.2 ± 3.8	6.2 ± 2.6
High-sensitivity C-reactive protein, mg/L	4.2 ± 8.6	0.9 ± 1.2	3.5 ± 7.8
Erythrocyte sedimentation rate, mm/hr	45.4 ± 35.9	27.0 ± 13.9	41.6 ± 33.2
MDA5 antibody titer, U/mL	214,843.8 ± 233,919.1	106,458.3 ± 101,082.3	193,166.7 ± 217,032.7
PON1 Q192R polymorphism	30% QQ, 52% QR, 17% RR	50% QQ, 33% QR, 17% RR	34% QQ, 48% QR, 17% RR
Paraoxonase activity, U/mL	546.4 ± 442.7	627.3 ± 511.5	562.6 ± 449.1
Lactonase activity, U/mL	19.0 ± 8.7	31.0 ± 26.1	21.4 ± 14.2
Arylesterase activity, U/mL	273.6 ± 161.8	279.6 ± 207.8	274.8 ± 168.0
Ferritin, ng/mL	5,995.1 ± 4,411.6	2,787.0 ± 3,099.4	5,353.5 ± 4,335.4
Initial medications[c]			
Glucocorticoids	21 (88)	6 (100)	27 (90)
Mycophenolate mofetil	8 (36)	2 (33)	10 (36)
Intravenous Ig	6 (27)	2 (33)	8 (29)
Rituximab	1 (5)	1 (17)	2 (7)
Cyclophosphamide (oral)	1 (5)	0 (0)	1 (4)
Tofacitinib	0 (0)	0 (0)	0 (0)
Tacrolimus	0 (0)	0 (0)	0 (0)
Hydroxychloroquine	5 (23)	2 (33)	7 (25)
Leflunomide	0 (0)	0 (0)	0 (0)
Azathioprine	0 (0)	0 (0)	0 (0)
Methotrexate	0 (0)	2 (33)	2 (7)
TNF inhibitor	0 (0)	0 (0)	0 (0)
Total number of therapies	1.9 ± 1.0	2.5 ± 1.4	2 ± 1.1
Time from diagnosis to medication date, mo	0.4 ± 2.1	−0.6 ± 5.7[d]	0.1 ± 3.1
Medications at last follow-up			
Glucocorticoids	15 (63)	2 (33)	17 (57)
Mycophenolate mofetil	16 (67)	4 (67)	20 (67)
Intravenous Ig	19 (79)	4 (67)	23 (77)
Rituximab	14 (58)	1 (17)	15 (50)
Cyclophosphamide (oral)	2 (8)	0 (0)	2 (7)
Tofacitinib	2 (8)	1 (17)	3 (10)
Tacrolimus	1 (4)	0 (0)	1 (3)
Hydroxychloroquine	3 (13)	0 (0)	3 (10)
Leflunomide	1 (4)	0 (0)	1 (3)
Azathioprine	0 (0)	0 (0)	0 (0)
Methotrexate	0 (0)	0 (0)	0 (0)

(Continued)

Table 1. (*Cont'd*)

	ILD (n = 24)	No ILD (n = 6)	Overall (n = 30)
TNF inhibitor	0 (0)	0 (0)	0 (0)
Total number of therapies	3.1 ± 1.3	2.0 ± 1.3	2.9 ± 1.3
Time from diagnosis to medication date, y	4.0 ± 3.7	6.7 ± 4.5	4.5 ± 4.0

* Values are mean ± SD or n (%). CPK, creatine phosphokinase; DM, dermatomyositis; ILD, interstitial lung disease; LDH, lactate dehydrogenase; PON1, paraoxonase 1; TNF, tumor necrosis factor; VAS, visual analog scale.
[a] Clinical characteristics at baseline cohort visit.
[b] Laboratory studies at baseline cohort visit.
[c] Initial medication use was not available by chart review for two patients in the ILD group.
[d] Mean time from diagnosis to medication date was negative in the no ILD group due to patients who started treatment for connective tissue disease before conclusive diagnosis of DM.

therapies was similar between groups. To evaluate the effects of time, group, and their interactions on the number of immunotherapies, the GEE method with a Poisson distribution was applied to account for within-subject correlation (Supplemental Table 2). The number of therapies gradually increased 1.34% (95% confidence interval [CI] 0.63%–2.05%, $P = 0.002$) per unit time. Patients in group A used 17.06% (95% CI 1.38%–35.18%, $P = 0.03$) more immunotherapies than those in group B. There's no strong evidence that the time trend differs by group ($P = 0.14$). Similar trends were obtained using mixed-effects

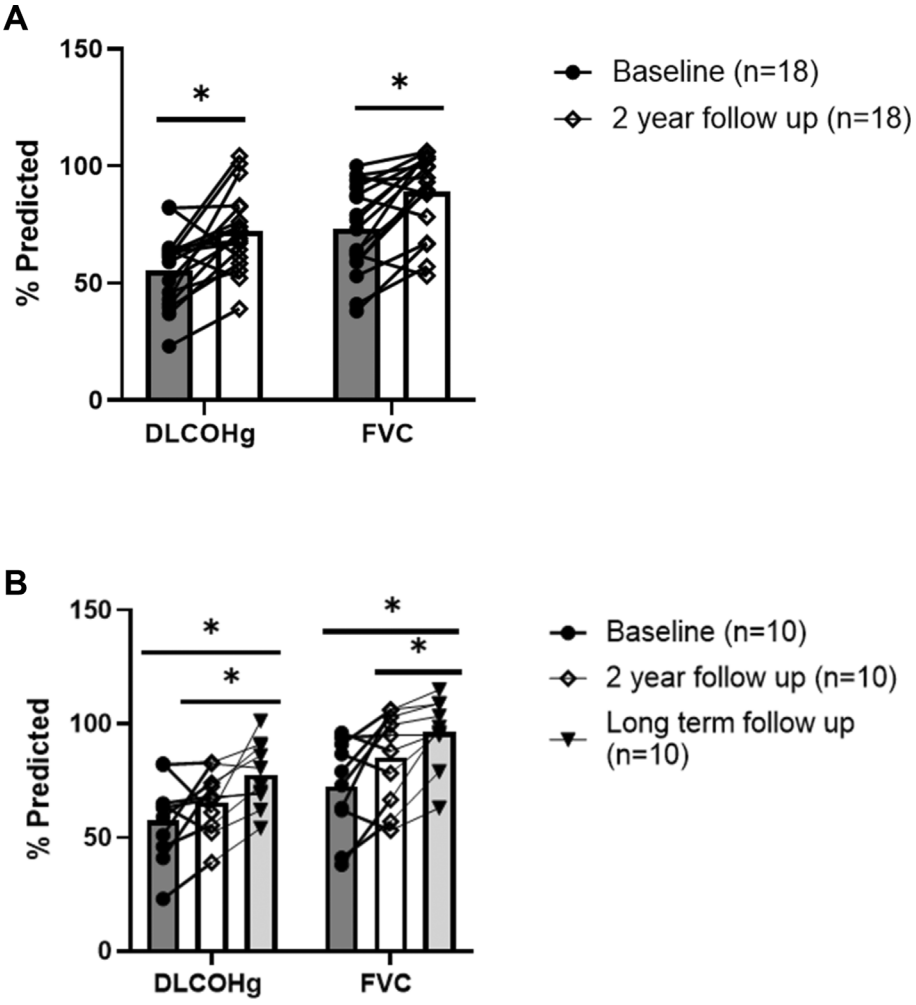

Figure 1. PFT results in patients with MDA5 antibody–positive DM-ILD. (A) PFT results of 18 patients with DM-ILD were available at approximately two years for review, specifically a mean ± SD of 1.8 ± 0.6 years. (B) PFT results of 10 patients with DM-ILD were also available at approximately seven years, specifically a mean ± SD of 6.8 ± 3.4 years. Data and error bars shown represent mean ± SEM. Individual data points are plotted over bar graphs. Paired *t*-tests were used for the comparison of PFT data between time points. *$P < 0.05$ for comparison between marked groups. DLco Hg, diffusing capacity for carbon monoxide adjusted for hemoglobin; DM, dermatomyositis; FVC, forced vital capacity; ILD, interstitial lung disease; PFT, pulmonary function test.

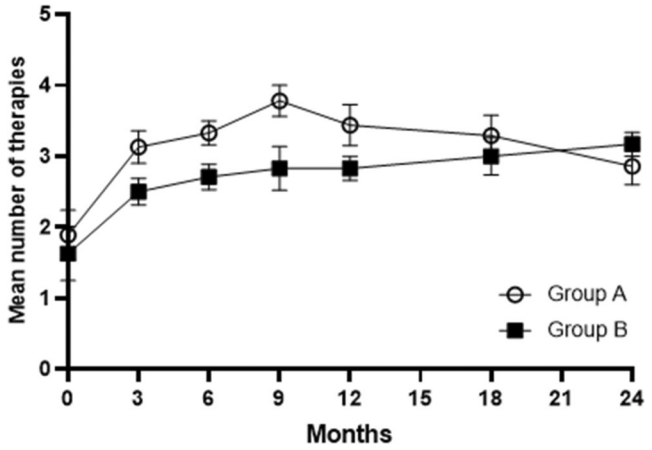

Figure 2. Mean number of therapies over time of patients with >20% relative improvement in FVC and DLco (group A, n = 10) versus patients with <20% relative improvement in FVC and DLco (group B, n = 7) at approximately two-year follow-up (mean ± SD 1.8 ± 0.6 years). Data were available for 17 patients. Mixed-effect models as well as the generalized estimating equations method with a Poisson distribution were used to examine the effects of time, group (≥20% vs <20% improvement in both FVC and DLco Hg), and their interaction on the number of immunotherapies. Detailed data are reported in Supplemental Table 2. Patients in group A with marked pulmonary response to treatment over longitudinal follow-up had a >40% percent increase in percent predicted FVC and DLco Hg from the baseline visit (n = 10; mean ± SD percent change 46% ± 5% FVC predicted, 54% ± 7% DLco Hg predicted). Patients in group B had less improvement in lung function compared to the baseline visit (n = 7; mean ± SD percent change 2% ± 4% FVC predicted, 17% ± 17% DLco Hg predicted). Of note, one patient in group B had a marked improvement in DLco Hg (126% percent change) but only a 4% change in FVC, leading to the large SD observed in the DLco Hg value. All data are represented as mean ± SEM. DLco Hg, diffusing capacity for carbon monoxide adjusted for hemoglobin; FVC, forced vital capacity.

model analysis. Adjustment for additional covariates in these models was not considered due to the small number of patients (n = 17); however, baseline demographics of the two groups were similar, with the exception for baseline percent predicted DLco and FVC, which was lower in group A (Supplemental Table 3).

Radiologic assessment of lung parenchymal changes over longitudinal follow-up. HRCT scans at baseline and the last follow-up were available for review in 21 of 24 patients with DM-ILD in the current cohort. Over longitudinal follow-up (mean ± SD 33 ± 31 months), there were significant decreases in extent of consolidations or ground glass opacities and trends for decreases in extent of small airways disease and bronchiectasis (Figure 3).

MDA5 antibody titers are associated with the presence of ILD and disease activity. All patients had a baseline blood specimen drawn upon entry into the cohort. Higher MDA5 antibody titers correlated significantly with higher overall disease activity, measured by physician global disease activity assessment per VAS at the same visit (correlation coefficient r = 0.49, P = 0.006). MDA5 antibody levels did not significantly correlate with the physician global assessment of damage VAS, muscle enzymes (CPK, lactate dehydrogenase [LDH], aldolase), or markers of inflammation (ESR, hsCRP) (data not shown). Patients with DM-ILD had a trend for higher MDA5 antibody levels at baseline compared to MDA5 antibody–positive patients without ILD (P = 0.10; Table 1).

In patients who had two-year follow-up PFT data available (n = 18; Figure 1A), numerical trends were noted for patients in group B who did not achieve 20% relative improvement in DLco and FVC over two-year longitudinal follow-up (n = 8) to have higher baseline MDA5 antibody titers (mean ± SD 243,052 ± 108,643 U/mL) compared to patients who achieved 20% improvement in DLco and FVC (n = 10) (mean ± SD 200,338 ± 289,054 U/mL; P = 0.12). Similar trends were not seen for other traditional disease measures at baseline,

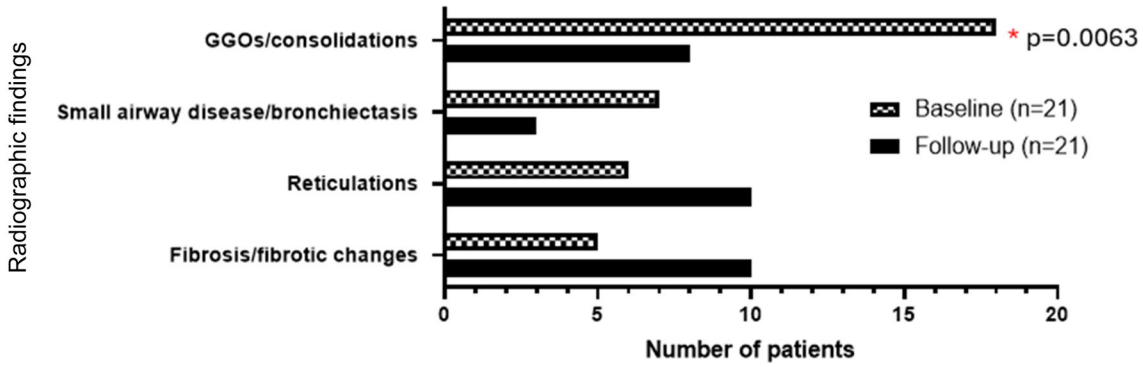

Figure 3. Comparison of DM-ILD patients' HRCT chest findings at baseline and follow-up. Of 24 patients with MDA5 antibody–positive DM-ILD in our cohort, 21 patients had baseline and interval follow-up HRCT chest scans. Mean ± SD time to interval follow-up HRCT chest scan was 33 ± 31 months for all patients. DM, dermatomyositis; GGO, ground glass opacities; HRCT, high-resolution computed tomography; ILD, interstitial lung disease. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80004/abstract.

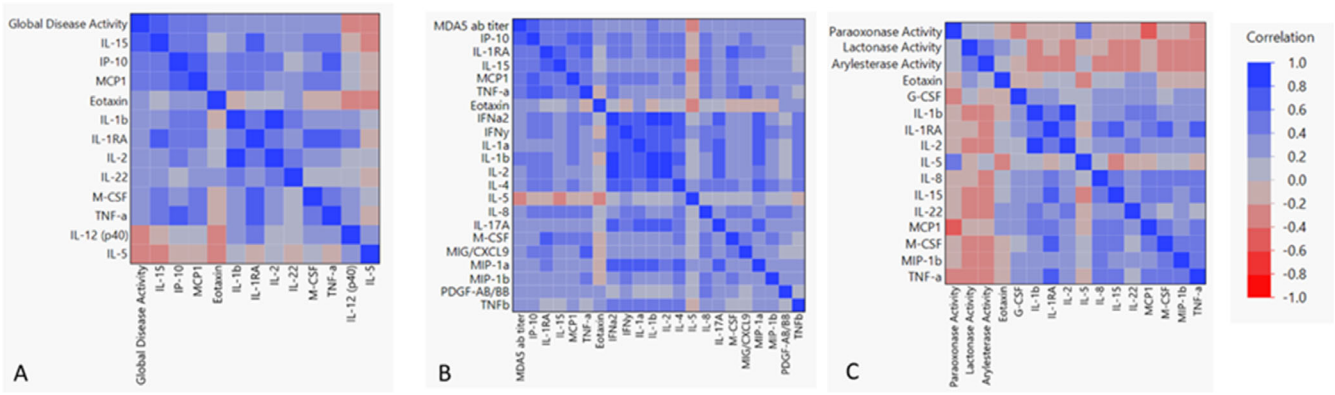

Figure 4. Heat maps of Spearman's ρ for correlations of circulating cytokines and chemokines with (A) physician global assessment of disease activity, (B) MDA5 antibody titers, and (C) paraoxonase 1 activities in longitudinal analyses.

including muscle enzymes, markers of inflammation, and physician global activity and damage assessments (Supplemental Table 3).

Longitudinal changes in disease activity, MDA5 antibody levels, and PON1 activity. Sixty-three percent (19/30, n = 4 DM, n = 15 DM-ILD) of patients in the cohort had longitudinal specimens in addition to baseline specimens available for analysis, which were collected at follow-up visits (DM: mean ± SD 2.3 ± 1.2 specimens per patient; DM-ILD: mean ± SD 2.2 ± 1.3 specimens per patient). In this expanded data set including longitudinal samples, lower MDA5 antibody titers maintained a strong correlation with lower overall disease activity measured by the physician global assessment of activity VAS (r = 0.67, $P < 0.0001$). In addition, lower MDA5 antibody levels were associated with greater enzyme activities of PON1, particularly its arylesterase activity (r = –0.40, $P = 0.0009$) and to a lesser degree its lactonase (r = –0.27, $P = 0.03$) and paraoxonase (r = –0.25, $P = 0.04$) activities.

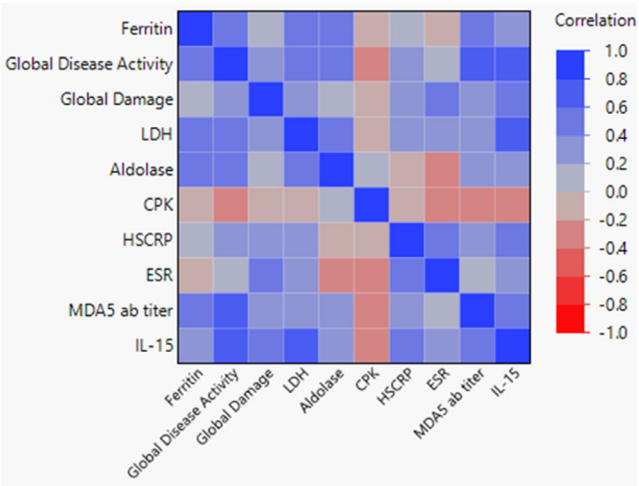

Figure 5. Heat map of Spearman's ρ for correlations of ferritin with other measures of disease activity and damage in longitudinal analyses. CPK, creatine phosphokinase; ESR, erythrocyte sedimentation rate; HSCRP, high-sensitivity C-reactive protein; IL, interleukin; LDH, lactate dehydrogenase.

Associations of circulating cytokines and chemokines with disease activity, MDA5 antibody titers, and PON1 activity. *Associations of cytokines and chemokines with disease activity and ILD.* Higher levels of IL-15 and lower levels of IL-12 (p40) correlated significantly with higher global disease activity, assessed by the physician global assessment VAS, at baseline (r = 0.52, $P = 0.003$, and r = –0.58, $P = 0.0008$, respectively). No other significant baseline associations of disease activity with other cytokines and chemokines were noted. IL-15 levels were also significantly higher in patients with DM-ILD (mean ± SD 30 ± 17 pg/mL) compared to patients with DM without ILD (mean ± SD 16 ± 9 pg/mL) ($P = 0.02$). No other cytokines or chemokines were significantly associated with the presence of ILD at baseline.

IL-15 remained strongly associated with global disease activity (r = 0.68, $P < 0.0001$) in expanded analyses of specimens collected at baseline as well as longitudinal follow-up (Figure 4A). IL-15 levels at baseline were also significantly higher in the two patients with DM-ILD who died compared to the rest of the cohort (mean ± SD 71.7 ± 0.4 pg/mL vs 24.1 ± 12.4 pg/mL; $P < 0.01$), and a trend was noted for lower lactonase activity in these two patients (mean ± SD 10.0 ± 5.5 U/mL vs 22.2 ± 14.3 U/mL, respectively; $P = 0.11$). IP-10 and MCP-1 levels were also moderately associated with disease activity in this analysis (r = 0.47, $P < 0.0001$, and r = 0.45, $P = 0.0002$, respectively). Additional correlations of cytokines and chemokines with global disease activity included eotaxin, IL-1β, IL-1RA, IL-2, IL-22, M-CSF, and TNFα (r range = 0.27–0.40, P range = 0.001–0.03) and IL-12 (p40) and IL-5 levels (r = –0.26, $P = 0.03$, and r = –0.40, $P = 0.001$, respectively).

Associations of cytokines and chemokines with MDA5 antibody titers. Higher circulating levels of IL-15 and IL-1Ra were

associated with higher MDA5 antibody levels at the baseline visit (r = 0.51, $P = 0.004$, and r = 0.58, $P = 0.0007$, respectively). Additional moderate correlations with MDA5 antibody were noted with IFNα2, IL-17A, IL-22, IP-10, MCP-1, and TNFα (r range = 0.37–0.46, P range = 0.03–0.04) and IL-5 levels (r = –0.44, $P = 0.01$). In the expanded longitudinal data set, similar associations were noted, with the strongest correlations observed between MDA5 antibody levels and IP-10 (r = 0.70), IL-1RA (r = 0.58), IL-15 (r = 0.57), MCP (r = 0.56), and TNFα (r = 0.51) (P for all < 0.0001) (Figure 4B). Additional positive but more moderate correlations in the expanded longitudinal data set with MDA5 antibody levels included circulating eotaxin, IFNα2, IFNγ, IL-1α, IL-1β, IL-2, IL-4, IL-5, IL-8, IL-17A, M-CSF, MIG/CXCL9, MIP-1α, MIP-1β, PDGF-AB/BB, and TNFβ (r range = 0.25–0.44, P range = 0.0004–0.04). A negative correlation remained with IL-5 levels and MDA5 antibody (r = –0.40, $P = 0.001$).

Associations of cytokines and chemokines with PON1 activities. Higher IL-15 levels were associated with lower lactonase activity of PON1 (r = –0.38, $P = 0.04$) at baseline. No other significant associations of cytokine or chemokine levels at baseline with PON1 activities were noted. In expanded analyses including longitudinal samples, the correlation of IL-15 with impaired lactonase activity of PON1 was maintained (r = –0.31, $P = 0.01$), and higher IL-15 levels were also associated with suppression of arylesterase activity of PON1 (r = –0.34, $P = 0.006$) (Figure 4C). Additional negative correlations of inflammatory cytokines and chemokines with lactonase and arylesterase activities of PON1 included TNFα, M-CSF, MIP-1β, and IL-2 (r range = –0.26 to –0.29, P range = 0.01–0.04 [lactonase]) and IL-22, IL-1β, IL-1RA, IL-8, M-CSF, MIP-1β, and TNFα (r range = –0.26 to –0.39, P range = 0.003–0.04 [arylesterase]). Higher concentrations of these cytokines and chemokines were associated with lower PON1 activities. A modest positive association of eotaxin with lactonase activity was noted (r = 0.25, $P = 0.04$). Finally, higher MCP-1 levels showed the strongest correlation with suppression of paraoxonase activity of PON1 in longitudinal analyses (r = –0.47, $P < 0.0001$), and additional modest correlations with paraoxonase activity occurred with TNFα and G-CSF (r = –0.28, $P = 0.02$ [both]) as well as with IL-5 levels (r = 0.40, $P = 0.0008$).

Association of ferritin levels with disease activity measures in MDA5 antibody–positive DM. Ferritin levels correlated significantly with physician global assessment of disease activity (r = 0.45, $P = 0.0002$) and MDA5 antibody (r = 0.47, $P < 0.0001$), IL-15 (r = 0.31,$P = 0.01$), LDH (r = 0.44, $P = 0.003$), and aldolase levels (r = 0.53, $P < 0.0001$) but did not correlate with physician global assessment of damage scales, CPK, ESR, or hsCRP (Figure 5).

DISCUSSION

This retrospective study examines the clinical course of patients with MDA5 antibody–positive DM from a single-center North American cohort. MDA5 antibody–positive DM is one of the most aggressive forms of DM due to its association with RP-ILD, which has a six-month mortality rate of 28% to 66% in some Asian cohort studies.[15,16] Less information is available regarding the pulmonary disease outcomes of North American populations with MDA5 antibody–positive DM, prompting the current investigation.

The major findings of the current work are the overall low mortality (n = 2/24 DM-ILD, 8%) and marked improvement in lung function noted in longitudinal follow-up of the patients with MDA5 antibody–positive DM-ILD, which is different from other connective tissue disease–associated ILDs (CTD-ILDs) such as rheumatoid arthritis and scleroderma-associated ILD. Seventy-five percent of the patients with DM-ILD had a history of hospitalization, attesting to the aggressive phenotype of the cohort. Two patients who died of rapidly progressive disease died early in the disease course, within seven months from diagnosis. In total, six patients had aggressive initial disease requiring supplemental oxygen on presentation, and four of those patients survived, three of whom no longer require oxygen therapy. In the 18 patients with DM-ILD with approximately two-year longitudinal PFT follow-up data available, there were absolute increases of 15% or greater in percent predicted DL_{CO} and FVC from baseline. In the 10 patients who had longer-term, approximately seven-year follow-up data at the time of the analysis, absolute increases of 20% or greater were noted in DL_{CO} and FVC from baseline. There was also significant improvement in lung function from the two-year to seven-year follow-up time points, suggesting that long-term improvement in lung function beyond the first two years of treatment is possible.

The observational nature of the current study and the relatively small number of patients are study limitations. However, the salient point from the current report is the feasibility of dramatic longitudinal improvement in lung function in patients with MDA5 antibody–positive DM-ILD, which may be different from other forms of CTD-ILD. In the Scleroderma Lung Study II, improvements of 2.19% and 2.88% in FVC predicted occurred after treatment with mycophenolate mofetil and cyclophosphamide, respectively.[17] These data reinforce the importance of a personalized, "precision medicine" approach to treatment of CTD-ILD, as recommended in the 2023 ACR/American College of Chest Physicians guidelines.[18]

The majority of patients with MDA5 antibody–positive DM-ILD in the current work were treated with combination therapy, with a mean of three therapies at the last follow-up visit, which was over four years from DM diagnosis. Patients with greater improvements on PFTs were treated with a higher number of therapies, suggesting that early aggressive therapy can be

associated with greater improvement in DM-ILD. Nearly 70% of patients at the last follow-up visit were receiving mycophenolate mofetil, ~60% were receiving rituximab, ~60% were receiving glucocorticoids (GC) (predominantly low dose), and nearly 80% of patients were receiving intravenous Ig (IVIg) therapy. Additional work by Wang et al has suggested potential benefits of IVIg on DM-ILD.[19]

Tsuji and colleagues first reported an association between aggressive early combination therapy and improved outcomes in 29 adult Japanese patients with new-onset MDA5 antibody–positive DM-ILD treated with high-dose GC, tacrolimus, and intravenous cyclophosphamide.[15] This combined immunosuppressive regimen group showed significantly higher six-month survival rates compared to a historical control group (n = 15) who received high-dose GC and stepwise addition of immunosuppressants (89% vs 33% respectively; $P < 0.0001$). In long-term follow-up at approximately 6.4 years in the "triple therapy" cohort, improvements in percent FVC and DLco predicted of 16% and 28%, respectively, were reported, similar to the changes reported in the current work.[16] Additional randomized controlled trials of different combination regimens in MDA5 antibody–positive DM-ILD may be of value.

We found significant correlations of MDA5 antibody titers with disease activity at baseline and longitudinally. A trend for patients with marked pulmonary response to therapy to have lower baseline MDA5 antibody titers was noted but not statistically significant. These data are consistent with prior small Asian studies suggesting associations of MDA5 antibody with disease activity and outcomes.[20] Accurate assessment of disease activity in patients with MDA5 antibody–positive DM is challenging because traditional laboratory assessments for DM, such as muscle enzyme assessments, may not correlate with disease activity. Here, mean CPK levels were in the normal range at the baseline visit despite a history of proximal muscle weakness in 88% of patients with DM-ILD and baseline moderate disease activity in patients with DM-ILD (mean physician VAS > 50). Therefore, identifying additional biomarkers that are associated with disease activity is important for appropriate treatment.

The current study investigated a large panel of cytokines and chemokines in the MDA5 antibody–positive DM cohort. Of the 37 analytes investigated, only IL-15 was significantly associated with the presence of DM-ILD at baseline. IL-15 was also the only cytokine to be associated with global disease activity and MDA5 antibody titers at baseline. A strong correlation persisted between higher IL-15 levels and higher global disease activity, even in our expanded longitudinal data set.

IL-15 is a cytokine critical to the development, survival, and activation of immune cells, including natural killer, T, and B cells.[21] IL-15 is expressed primarily within monocytes, macrophages, and dendritic cells and signals through the JAK/STAT pathway.[21–24] In a Japanese study of DM/polymyositis–associated ILD, IL-15 levels were significantly higher in patients with RP-ILD (n = 23) compared to patients without RP-ILD (n = 26), although anti-MDA5 antibody titers were similar between the groups.[25] Another Asian cohort also reported that serum IL-15 levels were significantly elevated in nine patients with MDA5 antibody–positive DM-ILD who did not survive compared to 17 patients who did survive.[26] Similarly, in the current work, IL-15 levels were significantly higher in the two patients with DM-ILD who died compared to the surviving patients. Further work assessing IL-15 as a disease activity biomarker in additional North American MDA5 antibody–positive cohorts may be of interest.

Finally, DM is a disease of the microvasculature, and patients with MDA5 antibody–positive DM exhibit marked disease manifestations resulting from vasculopathy.[27] Our previous work suggested that the function of HDL is abnormal in patients with IIM.[10] Specifically, PON1, the major antioxidant protein of HDL, which metabolizes oxidized phospholipids, was markedly impaired in patients with myositis with severe ILD and the presence of antisynthetase and MDA5 antibody.[11] Presently, we identify the activity of PON1 in our MDA5 antibody–positive DM cohort as one of the only circulating biomarkers besides IL-15 to be associated with disease activity at baseline and longitudinally. This may encourage further studies evaluating the role of PON1 and proinflammatory lipids in the pathogenesis of DM-ILD.

There are several limitations to the current observational study. The effects of specific treatments on disease activity, biomarkers, and pulmonary outcomes cannot be assessed due to the lack of a randomized controlled trial setting and the lack of prespecified time points for assessment. Another potential limitation of this study is survivorship bias, as the enrollment of newly and previously diagnosed patients may have led to a disproportionate representation of healthier patients with more favorable outcomes. Furthermore, comparisons of PFT data between group A and group B could not be adjusted for possible confounders due to the small sample size, including baseline PFT differences, which may have provided more room for improvement in group A. Future large prospective studies will also account for clustering of data in participants over longitudinal follow-up. All laboratory analyses must be considered pilot in nature because corrections for multiple comparisons were not performed due to the exploratory approach and small cohort size studied for this rare disease. Perhaps reassuring, however, are findings such as the disease activity association with IL-15, which is congruent with results of other Asian cohort studies.

In summary, this is one of the first North American studies to report marked improvement in pulmonary function in short- and long-term longitudinal follow-up in an MDA5 antibody–positive DM-ILD cohort. Similar to previous work in a trial cohort of 29 Japanese patients with MDA5 antibody–positive DM-ILD, favorable outcomes were identified with aggressive combination therapy, including 15%

to 20% improvements in percent predicted FVC and DL_{CO}.[15] Different from the Japanese study, the current cohort was not composed solely of patients with new-onset MDA5 antibody–positive DM-ILD, and patients were treated predominantly with mycophenolate mofetil, rituximab, GC, and IVIg over many years. Further trials are needed to examine different treatment approaches of combination therapy in MDA5 antibody–positive DM and ILD. Finally, the current work further supports a potential role for IL-15 and PON1 in the pathogenesis of MDA5 antibody–positive DM and ILD, which may warrant additional investigation.

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 Charles-Schoeman 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. Fuzzi E, Gatto M, Zen M, et al. Anti-MDA5 dermatomyositis: an update from bench to bedside. Curr Opin Rheumatol 2022;34(6):365–373. doi:https://doi.org/10.1097/BOR.0000000000000908
2. Love LA, Leff RL, Fraser DD, et al. A new approach to the classification of idiopathic inflammatory myopathy: myositis-specific autoantibodies define useful homogeneous patient groups. Medicine (Baltimore) 1991;70(6):360–374. doi:https://doi.org/10.1097/00005792-199111000-00002
3. Targoff IN. Idiopathic inflammatory myopathy: autoantibody update. Curr Rheumatol Rep 2002;4(5):434–441. doi:https://doi.org/10.1007/s11926-002-0089-7
4. Mehta P, Machado PM, Gupta L. Understanding and managing anti-MDA 5 dermatomyositis, including potential COVID-19 mimicry. Rheumatol Int 2021;41(6):1021–1036. doi:https://doi.org/10.1007/s00296-021-04819-1
5. Hall JC, Casciola-Rosen L, Samedy LA, et al. Anti-melanoma differentiation-associated protein 5-associated dermatomyositis: expanding the clinical spectrum. Arthritis Care Res (Hoboken) 2013;65(8):1307–1315. doi:https://doi.org/10.1002/acr.21992
6. Fiorentino D, Chung L, Zwerner J, et al. The mucocutaneous and systemic phenotype of dermatomyositis patients with antibodies to MDA5 (CADM-140): a retrospective study. J Am Acad Dermatol 2011;65(1):25–34. doi:https://doi.org/10.1016/j.jaad.2010.09.016
7. Takada T, Aoki A, Asakawa K, et al. Serum cytokine profiles of patients with interstitial lung disease associated with anti-CADM-140/MDA5 antibody positive amyopathic dermatomyositis. Respir Med 2015;109(9):1174–1180. doi:https://doi.org/10.1016/j.rmed.2015.07.004
8. Hallowell RW, Danoff SK. Diagnosis and management of myositis-associated lung disease. Chest 2023;163(6):1476–1491. doi:https://doi.org/10.1016/j.chest.2023.01.031
9. Linton MF, Yancey PG, Tao H, et al. HDL Function and atherosclerosis: reactive dicarbonyls as promising targets of therapy. Circ Res 2023;132(11):1521–1545. doi:https://doi.org/10.1161/CIRCRESAHA.123.321563
10. Bae SS, Lee YY, Shahbazian A, et al. High- density lipoprotein function is abnormal in idiopathic inflammatory myopathies. Rheumatology (Oxford) 2020;59(11):3515–3525. doi:https://doi.org/10.1093/rheumatology/keaa273
11. Bae SS, Shahbazian A, Wang J, et al. Abnormal paraoxonase-1 (PON1) enzyme activity in idiopathic inflammatory myopathies. Rheumatology (Oxford) 2022;61(6):2512–2523. doi:https://doi.org/10.1093/rheumatology/keab795
12. Lundberg IE, Tjärnlund A, Bottai M, et al. 2017 European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Arthritis Rheumatol 2017;69(12):2271–2282.
13. Charles-Schoeman C, Wang J, Shahbazian A, et al. Suppression of inflammatory arthritis in human serum paraoxonase 1 transgenic mice. Sci Rep 2020;10(1):16848. doi:https://doi.org/10.1038/s41598-020-74016-w
14. Charles-Schoeman C, Lee YY, Shahbazian A, et al. Association of paraoxonase 1 gene polymorphism and enzyme activity with carotid plaque in rheumatoid arthritis. Arthritis Rheum 2013;65(11):2765–2772. doi:https://doi.org/10.1002/art.38118
15. Tsuji H, Nakashima R, Hosono Y, et al. Multicenter prospective study of the efficacy and safety of combined immunosuppressive therapy with high-dose glucocorticoid, tacrolimus, and cyclophosphamide in interstitial lung diseases accompanied by anti-melanoma differentiation-associated gene 5-positive dermatomyositis. Arthritis Rheumatol 2020;72(3):488–498. doi:https://doi.org/10.1002/art.41105
16. Sasai T, Nakashima R, Tsuji H, et al. Long-term prognosis of antimelanoma differentiation-associated gene 5-positive dermatomyositis with interstitial lung disease. J Rheumatol 2023;50(11):1454–1461. doi:https://doi.org/10.3899/jrheum.2023-0371
17. Tashkin DP, Roth MD, Clements PJ, et al; Sclerodema 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. doi:https://doi.org/10.1016/S2213-2600(16)30152-7
18. Johnson SR, Bernstein EJ, Bolster MB, et al. 2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) guideline for the treatment of interstitial lung disease in people with systemic autoimmune rheumatic diseases. Arthritis Rheumatol 2024;76(8):1182–1200. doi:https://doi.org/10.1002/art.42861
19. Wang LM, Yang QH, Zhang L, et al. Intravenous immunoglobulin for interstitial lung diseases of anti-melanoma differentiation-associated gene 5-positive dermatomyositis. Rheumatology (Oxford) 2022;61(9):3704–3710. doi:https://doi.org/10.1093/rheumatology/keab928
20. Sato S, Kuwana M, Fujita T, et al. Anti-CADM-140/MDA5 autoantibody titer correlates with disease activity and predicts disease outcome in patients with dermatomyositis and rapidly progressive interstitial lung disease. Mod Rheumatol 2013;23(3):496–502. doi:https://doi.org/10.3109/s10165-012-0663-4
21. Mishra A, Sullivan L, Caligiuri MA. Molecular pathways: interleukin-15 signaling in health and in cancer. Clin Cancer Res 2014;20(8):2044–2050. doi:https://doi.org/10.1158/1078-0432.CCR-12-3603
22. Grabstein KH, Eisenman J, Shanebeck K, et al. Cloning of a T cell growth factor that interacts with the beta chain of the interleukin-2 receptor. Science 1994;264(5161):965–968. doi:https://doi.org/10.1126/science.8178155
23. Bamford RN, Battiata AP, Burton JD, et al. Interleukin (IL) 15/IL-T production by the adult T-cell leukemia cell line HuT-102 is associated with a human T-cell lymphotrophic virus type I region /IL-15 fusion message that lacks many upstream AUGs that normally attenuates

IL-15 mRNA translation. Proc Natl Acad Sci U S A 1996;93(7):2897–2902. doi:https://doi.org/10.1073/pnas.93.7.2897

24. Blauvelt A, Asada H, Klaus-Kovtun V, et al. Interleukin-15 mRNA is expressed by human keratinocytes Langerhans cells, and blood-derived dendritic cells and is downregulated by ultraviolet B radiation. J Invest Dermatol 1996;106(5):1047–1052. doi:https://doi.org/10.1111/1523-1747.ep12338641
25. Shimizu T, Koga T, Furukawa K, et al. IL-15 is a biomarker involved in the development of rapidly progressive interstitial lung disease complicated with polymyositis/dermatomyositis. J Intern Med 2021;289(2):206–220. doi:https://doi.org/10.1111/joim.13154
26. Takada T, Ohashi K, Hayashi M, et al. Role of IL-15 in interstitial lung diseases in amyopathic dermatomyositis with anti-MDA-5 antibody. Respir Med 2018;141:7–13. doi:https://doi.org/10.1016/j.rmed.2018.06.012
27. Kurtzman DJB, Vleugels RA. Anti-melanoma differentiation-associated gene 5 (MDA5) dermatomyositis: a concise review with an emphasis on distinctive clinical features. J Am Acad Dermatol 2018;78(4):776–785. doi:https://doi.org/10.1016/j.jaad.2017.12.010

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1123–1129
DOI 10.1002/acr.80015
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

American College of Rheumatology
Empowering Rheumatology Professionals

Immune Checkpoint Inhibitor–Related Myositis and Associated Triad Overlap Syndrome

Selene Rubino, MD,[1,2] Grant W. Cannon, MD,[2,3] Brian C. Sauer, PhD,[3] Jorge Rojas, MS,[3,4] Gregory J. Stoddard, MBA, MPH,[2] Gary A. Kunkel, MD,[2,3] Jessica A. Walsh, MD,[2,3] Bryant R. England, MD, PhD,[5,6] Joshua F. Baker, MD, MSCE,[7,8] and Tawnie J. Braaten, MD[2,3]

Objective. Immune checkpoint inhibitor (ICI) myositis is a rare but a highly morbid condition, particularly with the ICI myositis triad syndrome of myositis, myocarditis, and myasthenia gravis. We report the clinical characteristics of ICI myositis and all-cause mortality in these patients.

Methods. We performed a national retrospective cohort study of US veterans treated with an ICI between June 2011 and February 2022. Cases were identified by elevated creatinine kinase (CK) following ICI therapy and/or "myositis" in notes. Chart review determined if candidates had ICI-associated myositis and were subsequently classified as isolated ICI myositis or ICI triad syndrome.

Results. Of 29,296 veterans treated with ICIs, 68 (0.2%) ICI-induced myositis cases were identified, including 23 patients with ICI triad syndrome and 45 patients with isolated ICI myositis. Median survival duration for ICI triad syndrome was 2.5 (95% confidence interval [CI] 1.7–18.5) months in comparison to 20.4 (95% CI 6.2–60) months for isolated ICI myositis and 12.6 (95% CI 12.3–12.8) months median survival for patients treated with ICI without myositis. Patients with ICI triad syndrome had 60.0% mortality during the six months following initiation of ICI treatment, in comparison to 30.2% mortality in isolated ICI myositis and 31.5% mortality in patients without myositis. Six months after the first ICI treatment, subsequent mortality trends were similar in the three groups.

Conclusion. ICI triad syndrome is associated with higher overall mortality and shortened median survival times during the first six months after initial ICI therapy, whereas isolated ICI myositis is associated with similar mortality outcomes compared to patients treated with ICI who did not develop myositis.

INTRODUCTION

Since the introduction of the first immune checkpoint inhibitor (ICI), ipilimumab, in 2011, there has been a considerable expansion in number of ICIs and indications for ICIs.[1] The number of adverse events of ICIs have also increased during the expanded use of these medications. Although myositis is a rare complication of ICI therapy, representing 0.6% of all immune-related adverse events (irAEs) in pharmacovigilance databases,[2,3] there is considerable morbidity and significant mortality associated with its development. In 2020, Aldrich et al reported increased mortality in a subgroup of patients with ICI myositis who developed myocarditis and myasthenia gravis in addition to myositis following immunotherapy,[4] termed ICI myositis triad syndrome. In line with findings of Aldrich et al, subsequent retrospective cohort studies measured the overall mortality of ICI myositis triad to be between 42% and 51%.[5,6] Among patients with ICI myositis who died, those who had overlapping myocarditis or myasthenia gravis–like features were more likely to die of myositis irAE than of the progression of cancer.[7] The purpose of the current study was to evaluate the clinical features and outcomes of patients within the Veterans Health Administration (VHA) who developed ICI-induced myositis and ICI myositis triad compared to patients treated with ICI who did not develop myositis.

PATIENTS AND METHODS

Study design and data resources. We performed a national retrospective cohort study of veterans enrolled in VHA

[1]Mayo Clinic, Rochester, Minnesota; [2]University of Utah, Salt Lake City; [3]VA Salt Lake City Health Care System, Salt Lake City, Utah; [4]VA Puget Sound Health Care System, Seattle, Washington; [5]VA Nebraska-Western Iowa Health Care System, Omaha, Nebraska; [6]University of Nebraska Medical Center, Omaha; [7]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania; [8]University of Pennsylvania, Philadelphia.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80015.

Address correspondence via email to Selene Rubino, MD, at Rubino.Selene@mayo.edu.

Submitted for publication June 19, 2025; accepted in revised form January 22, 2026.

SIGNIFICANCE & INNOVATIONS

- Two immune checkpoint inhibitor (ICI)–associated myositis phenotypes were observed: ICI myositis triad syndrome and isolated ICI myositis.
- ICI myositis triad syndrome is typically a severe acute illness with significant risk for death shortly after initiation of ICI treatment.
- Isolated ICI myositis is a less severe disease with less mortality than ICI myositis triad syndrome with outcomes generally similar to patients treated with ICI who do not develop myositis.

who underwent ICI therapy between June 6, 2011, and February 14, 2023. Electronic health record (EHR) data from the VA Corporate Data Warehouse, including demographic data, clinical notes, pharmacy data, and laboratory results, were extracted and analyzed in these patients.[8] The VA Central Cancer Registry data (Oncology Raw Domain) provided information on cancer diagnoses, and mortality was obtained from the VHA Death Ascertainment File.[9]

Case identification and classification. Briefly, patients were selected for chart review if they had one of the following:

1. A creatinine kinase (CK) level raised above 10 times the upper limit of normal (ULN) within six months of ICI administration
2. A CK level raised above 1 × ULN within six months of ICI administration and whose EHR included the term "myositis" in any discharge summary after first ICI administration
3. A CK level raised above 1 × ULN and the term "myositis" within progress notes within six months after the first ICI treatment

Our process for identifying patients with ICI myositis is presented in Figure 1. All patients treated with ICI with CK measured within six months of first ICI infusion were identified. We then searched for the text term "myositis" in clinic notes and discharge summaries, identifying patients with high levels of CK and *International Classification of Diseases* (ICD) codes for idiopathic inflammatory myositis. The review of text notes involved two levels of review. The level one review was conducted by GWC and involved reading notes to determine if a specific explanation for the "myositis" term other than ICI myositis was present. The majority of these conditions included statin-induced myositis, baseline prior inflammatory myositis, the myositis term included in a problem list without evidence of any active muscle disorder, and negation statements such as "myositis absent." Cases considered to potentially be ICI myositis were subjected to level two review by SR. This level two review assessed the entire medical record in addition to clinical notes and included assessment of consultations, imaging, and laboratory testing to determine if the patient met the ICI myositis criteria listed below. Cases in which there were any concerns about the diagnosis were adjudicated by SR and TJB.

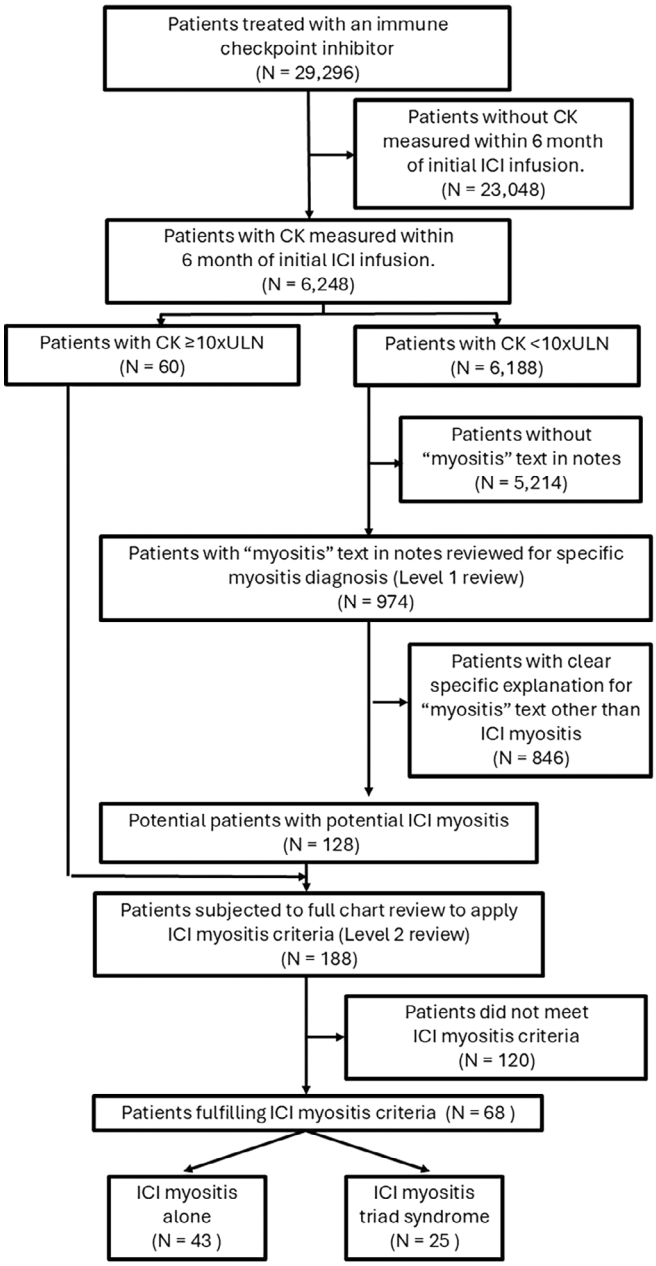

Figure 1. Flowchart outlining study selection and case categorization. CK, creatinine kinase; ICI, immune checkpoint inhibitor; ULN, upper limit of normal.

In addition to the review of all patients with measured CK and "myositis" text in clinical notes, all patients with one or more CK level ≥10 × ULN were subjected to level one review. A preliminary review of patients with CK levels elevated between 2 × and 10 × ULN and CK levels elevated above 10 × ULN showed ICI myositis in 30% (3 of 10) and 42% (8 of 19) of patients, respectively. The 20 patients with "myositis" text notes and CK levels

>10 × ULN were subjected to both level one and level two review. Of these 20 patients with both CK levels >10 × ULN and "myositis" text in clinic notes, 19 (95%) were classified as having ICI myositis on level two review. Thus, 10 × ULN was a value at which individual chart review should be considered for all patients with this level of CK elevation. ICD codes for idiopathic inflammatory myositis were associated with note text and not found to be useful for identifying ICI myositis. Of the 25 cases with idiopathic inflammatory myositis codes, 20 (80%) were identified by the screening of "myositis" text in notes, and a review of the other 5 cases did not show ICI myositis.

Patients were classified as having ICI-induced myositis if they had: (1) a confirmed provider diagnosis of ICI-induced myositis in the medical record by chart review, (2) an elevated CK level above the ULN after ICI therapy, and (3) no other explanation for the elevated CK level other than ICI-induced myositis. Patients meeting ICI-induced myositis criteria were further classified as ICI triad or isolated ICI myositis. The category "ICI triad syndrome" denotes patients who developed myositis, myasthenia gravis, and myocarditis at any point after the ICI was first administered. Myocarditis was defined as "possible" or "probable" myocarditis using the criteria proposed by Bonaca et al.[10] Myasthenia gravis was determined by symptoms of ptosis, diplopia, and/or respiratory failure with associated decline in negative inspiratory force (dysphagia was excluded due to association with myositis). Patients with ICI myositis on chart review who did not develop all three manifestations of ICI triad were classified as "isolated ICI myositis," including patients with overlap myocarditis and myositis without myasthenia gravis. All other patients were classified as "not ICI myositis."

Demographic information collected included age, sex, self-identified race, smoking status, and cancer diagnoses at the time of first date of ICI administration, which was defined as the index date. Individual ICI medications were recorded, and each agent further was classified as PD-1, PD-L1, and CTLA-4 inhibitors. Combination ICI therapy was defined as the delivery of both ipilimumab and nivolumab or durvalumab and tremelimumab concurrently.

Statistical analyses. For Table 1, the continuous variable age was compared using analysis of variance and the categorical variables compared using the chi-square test. The primary outcome was all-cause mortality. Crude mortality was calculated as the percentage of patients surviving from the time of the initial ICI infusion to six months, one year, and two years, respectively. Confidence intervals (CIs) for crude mortality were calculated using a binomial calculator and, for the median survival, using a Kaplan–Meier estimator. The median survival for each myositis group was the median duration above which an equal number of patients had died and below which a similar number had survived with longevity greater than the median. Median survival was compared using a log rank test. Crude hazard ratios and hazard ratios adjusted for age, race, smoking status, cancer type, and ICI class were determined from the Cox regression models from the time of initial ICI infusion to six-month, one-year, two-year, and five-year follow-up. For the Cox regression models and the Kaplan–Meier survival analysis, patients were censored at five years if surviving five years after the initial ICI infusions. To reduce potential of overfitting the model, the variables used in the adjustment were evaluated as dichotomous variables: age, over 70 years; race, White versus non-White; smoking status, ever versus never/unknown; cancer type, lung versus non-lung; and ICI class PD-1 versus other ICI classes. Sex was not included in the model because all patients with ICI myositis were male. Statistical analyses were performed using STATA version 18 (StataCorp, LLC, College Station, Texas).

Regulatory compliance. This study was approved by the University of Utah Institutional Review Board (IRB), which also serves as the IRB for the VA Salt Lake City Health Care System.

RESULTS

There were 29,296 patients in VHA treated with an ICI during the study period. Of these, 6,248 (21%) had a CK level measured within six months of ICI infusion (Figure 1). There were 974 patients with myositis documented in the EHR. After exclusion of patients with clear evidence of conditions other than ICI myositis and inclusion of all patients with 10 × ULN CK levels, there were 188 patients who underwent level two chart review. There were 120 patients who did not meet criteria for ICI myositis. The most common conditions seen in patients undergoing level two chart review who did not have ICI myositis were septic shock, cardiogenic shock, myocardial infraction, myositis associated with infections, and rhabdomyolysis. After medical record review, 68 patients (0.2%) were categorized as ICI myositis. Of the patients with ICI myositis, 43 patients were further categorized as isolated ICI myositis and 25 patients as ICI triad.

Baseline characteristics. Patients with ICI triad were significantly older, with a mean age at first ICI infusion of 76.0 years versus 70.8 years in patients with isolated ICI myositis and 69.9 years in patients without myositis ($P = 0.0013$; Table 1). All patients with ICI myositis (both ICI triad and isolated ICI myositis) were male. Patients with ICI triad were more likely to be White than patients with isolated ICI myositis (100% vs 67.4%, $P = 0.021$). No significant differences were found between smoking statuses. There were no differences in demographics between patients with isolated ICI myositis and patients treated with ICI who did not develop myositis.

ICI treatment. The initial ICIs prescribed were atezolizumab, avelumab, cemiplimab, durvalumab, ipilimumab, nivolumab, pembrolizumab, and tremelimumab (Table 1). The most common ICI class was PD-1 inhibitors, composing 73.7% of total ICIs prescribed, followed by PD-L1 inhibitors, which made up 20.5% of all

Table 1. Baseline characteristics at time of first ICI infusion*

	ICI triad syndrome (n = 25)	Isolated ICI myositis (n = 43)	No ICI myositis (n = 29,228)	Three group comparison
Age at initial ICI infusion, mean ± SD, y	76.0 ± 7.3	70.8 ± 9.9	69.9 ± 8.5	0.0013
Male sex, n (%)	25 (100)	43 (100)	28,272 (96.7)	0.317
Race, n (%)				0.021
Black or African American	0 (0)	13 (30.2)	5,187 (17.8)	
White	25 (100)	29 (67.4)	22,934 (78.5)	
Other	0 (0)	1 (2.3)	1,107 (3.8)	
Smoking status, n (%)				0.332
Ever	12 (48)	19 (44.2)	15,875 (53.8)	
Never	4 (16)	10 (23.3)	4,000 (13.6)	
Unknown	9 (36)	14 (32.6)	9,618 (32.6)	
Cancer diagnosis, n (%)				0.0001
Gastrointestinal	2 (8.0)	5 (11.6)	3,133 (10.7)	
Genitourinary	11 (44.0)	7 (16.3)	3,996 (13.7)	
Lung	3 (12.0)	20 (46.5)	13,349 (45.7)	
Melanoma	4 (16.0)	7 (16.3)	2,578 (8.8)	
Other	5 (20.0)	4 (9.3)	6,171 (21.1)	
Initial ICI therapy, n (%)				0.443
ICI monotherapy				
PD-1				
Cemiplimab	0 (0)	1 (2.3)	265 (0.9)	
Nivolumab	9 (36.0)	13 (30.2)	8,728 (29.9)	
Pembrolizumab	12 (48)	22 (51.2)	12,544 (44.0)	
Subtotal	21 (84)	36 (83.7)	21,537 (73.7)	
PD-L1				
Atezolizumab	0 (0)	1 (2.3)	3,078 (10.8)	
Avelumab	1 (4.0)	1 (2.3)	135 (0.5)	
Durvalumab	1 (4.0)	3 (7.0)	2,784 (9.8)	
Subtotal	2 (8.0)	6 (14)	5,997 (20.5)	
CTLA-4				
Ipilimumab	2 (8)	2 (4.7)	992 (3.4)	
Subtotal	2 (8)	2 (4.7)	992 (3.4)	
ICI combination therapy[a]				
Ipilimumab + nivolumab	0 (0)	0 (0)	699 (2.4)	
Durvalumab + tremelimumab	0 (0)	0 (0)	3 (0.01)	
Subtotal	0 (0)	0 (0)	702 (2.4)	

* ICI, immune checkpoint inhibitor.
[a] Combination therapy: the simultaneous delivery of ipilimumab and nivolumab or durvalumab and tremelimumab at first ICI infusion.

ICIs (Table 1). There were no significant differences between cancer immunotherapy class or drug and ICI myositis status. Combination therapy, defined as concurrent administration of two different ICIs at time of initial therapy, was not the initial therapy in any patients who developed isolated ICI myositis or ICI triad syndrome.

Cancer history. The most common cancers were lung, genitourinary, melanoma, and gastrointestinal. Patients with ICI triad had more genitourinary cancers (44.0% vs 16.3%) and fewer lung cancers (12.0% vs 46.5%) when compared to patients with isolated ICI myositis ($P = 0.0001$).

Laboratory outcomes. Patients with ICI triad had a significantly higher and earlier onset of CK elevation when compared to isolated ICI myositis. The first point that CK was measured after ICI therapy in patients with ICI triad was earlier than in patients with isolated myositis (29.3 vs 49.3 days, $P = 0.005$). The initial CK and CK-MB levels measured following ICI therapy were significantly higher in patients with ICI triad (5,064 vs 3,457 IU/L, $P = 0.0001$, and 152 vs 110 IU/L, $P = 0.0001$, respectively). The time to peak CK measurement was significantly shorter in patients with ICI triad compared to isolated ICI myositis at 29.1 days versus 53.0 days ($P = 0.04$). Similarly, the peak CK level attained after ICI was delivered was higher in patients with ICI triad (5,204 vs 4,416, $P = 0.0001$). Myositis antibodies were rarely ordered and largely negative among patients with ICI myositis.

Skeletal muscle biopsy was obtained in three patients with ICI triad. Two of the three demonstrated acute necrotizing myopathy. Immunostaining revealed a few endomysial T cells, no significant B cells, and prominently perimysial with scattered

Table 2. All-cause mortality following first ICI*

	ICI triad syndrome (n = 25)	Isolated ICI myositis (n = 43)	No ICI myositis (n = 29,228)	Three group comparison[a]
6-mo all-cause mortality, % (95% CI)	60 (38.7–78.9)	30.2 (17.2–46.1)	31.5 (31.0–32.0)	0.013
1-y all-cause mortality, % (95% CI)	64 (42.5–82.0)	41.9 (27.0–57.9)	48.9 (48.3–49.4)	0.216
2-y all-cause mortality, % (95% CI)	72 (50.6–87.9)	53.5 (37.7–68.8)	65.8 (65.3–66.4)	0.211
Median 50% survival, months (95% CI)[b]	2.5 (1.7–18.5)	20.4 (6.2–60.0)	12.6 (12.3–12.9)	0.1651

* CI, confidence interval; ICI, immune checkpoint inhibitor; IQR, interquartile range.
[a] Three group comparisons by chi-square or all-cause mortality and log rank test for median survival.
[b] Median months at 50% mortality and 95% CIs.

endomysial histiocytes. The third showed no evidence of inflammatory or toxic myopathy.

Mortality outcomes. During the six months after initial ICI infusion, patients developing ICI triad syndrome had median 60% (95% CI 38.72%–78.96%) mortality, which was higher than the 30.2% (17.2%–46.1%) mortality seen in patients developing isolated ICI myositis and 31.5% (31.0%–32.0%) mortality in patients without ICI myositis (Tables 2 and 3). Mortality was also higher in the ICI triad group at one year and two years of observation in comparison to the isolated ICI myositis and no ICI myositis groups. These differences were not statistically significant; however, interpretation is limited due to the increasingly low sample sizes at each time point. The median survival time was 2.5 (1.7–18.5) months in patients with ICI triad syndrome in comparison to the median survival of 20.4 (6.2–60) months in patients with isolated ICI myositis and 12.6 (12.3–12.9) months in patients without myositis.

The Kaplan–Meier survival curve for patients with ICI triad, patients with isolated ICI myositis, and patients without myositis over the five years following initial ICI infusion shows a high mortality in patients with ICI triad during the first year with a subsequent merging of mortality of the three groups over the following years (Figure 2). During the first six months after initial ICI infusion, Cox regression models showed higher hazard ratios for the patients with ICI triad group in comparison to patients with isolated ICI myositis and without myositis at 2.37 (95% CI 1.12–5.00) and 2.61 (95% CI 1.57–4.33) (Tables 2 and 3). After one year of follow-up, the mortality of the ICI triad group remained high in comparison to isolated ICI myositis; however, results were not statistically significant. After adjusting the Cox regression model for age, race, smoking status, cancer group, and initial ICI class, the hazard ratios over six months comparing patients with ICI triad to patients with isolated ICI myositis and no ICI myositis were 1.75 (95% CI 0.77–3.98) and 2.51 (95% CI 1.51–4.16). Because all patients with ICI myositis were male, the Cox regressions could not be adjusted for sex; however, an analysis limited to only male patients in all groups showed similar results to the full population (data not shown).

DISCUSSION

This report is the first nationwide multicenter cohort study evaluating ICI-associated myositis outcomes and represents one of the largest studies to date on ICI myositis triad. The size of our cohort afforded us a unique opportunity to compare ICI-

Table 3. All-cause mortality hazard ratios following first immune checkpoint inhibitor*

	ICI triad vs isolated ICI myositis	ICI triad vs no ICI myositis	Isolated ICI myositis vs no ICI myositis	ICI triad vs isolated ICI myositis[a] *P* value	ICI triad vs no ICI myositis[a] *P* value	Isolated ICI myositis vs no ICI myositis[a] *P* value
Hazard ratio—Unadjusted (95% CI)						
6 mo	2.37 (1.12–5.00)	2.61 (1.57–4.33)	0.98 (0.57–1.70)	0.023	0.0001	0.957
1 y	1.78 (0.98–3.80)	1.86 (1.14–3.03)	0.85 (0.54–1.35)	0.057	0.013	0.492
2 y	1.78 (0.96–3.32)	1.52 (0.96–2.41)	0.76 (0.50–1.14)	0.068	0.076	0.183
5 y	1.51 (0.86–2.65)	1.29 (0.83–2.00)	0.66 (0.38–1.17)	0.156	0.256	0.156
Hazard ratio—adjusted[b] (95% CI)						
6 mo	1.75 (0.77–3.98)	2.51 (1.51–4.16)	0.97 (0.56–1.66)	0.186	0.0001	0.904
1 y	1.79 (1.10–2.92)	1.79 (1.10–2.92)	0.84 (0.53–1.33)	0.020	0.020	0.46
2 y	1.46 (0.92–2.33)	1.47 (0.92–2.33)	0.75 (0.50–1.13)	0.104	0.104	0.166
5 y	1.20 (0.62–2.32)	1.22 (0.77–1.91)	0.74 (0.51–1.06)	0.588	0.396	0.100

* CI, confidence interval; ICI, immune checkpoint inhibitor.
[a] Three group comparisons by chi-square or all-cause mortality and log rank test for median survival.
[b] Adjusted for age, race, smoking status, cancer type, and class of initial ICI infusion.

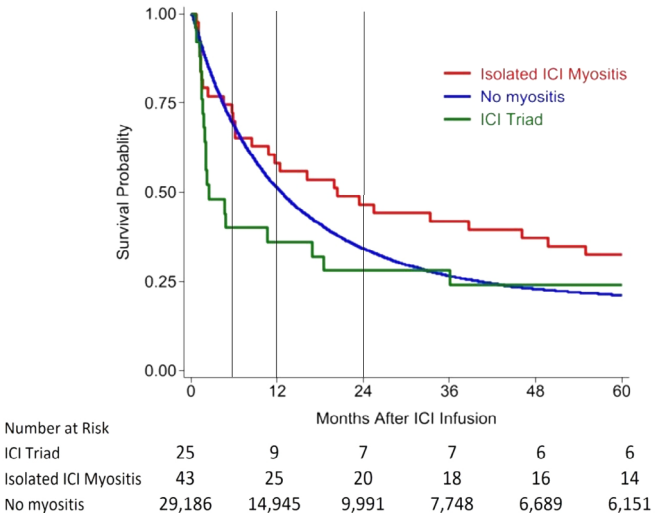

Figure 2. Kaplan–Meier curve demonstrating increased mortality of patients with ICI triad. ICI, immune checkpoint inhibitor.

associated myositis to the general population treated with ICI. These data confirm that although ICI-associated myositis is a significant irAE, the highest mortality is in patients with ICI myositis triad syndrome. We found that patients with ICI myositis triad had substantially increased mortality during the six months after the first infusion, though it is unclear if there is an increased risk for those that survive past this first six-month period.

Although patients with ICI myositis triad in our cohort had a higher mortality than patients without ICI myositis, patients with isolated ICI myositis had a similar survival compared to patients who did not develop myositis. This result may be partially due to the small sample size in the isolated ICI myositis group. Alternately, Saygin et al asserted that, in patients with ICI myositis, the most common cause of death is cancer progression; in overlap patients, it is the irAE.[11] If true, that would suggest that patients who survive ICI myositis may experience improved cancer outcomes. Although we were not able to capture progression of disease in our study, our results support previous findings that the cancer response rate may be higher in patients with irAEs.[12–14]

This study also demonstrated the early onset of ICI myositis triad and isolated ICI myositis after ICI initiation with the time to peak CK measurement of 29.1 days and 53.0 days, respectively. This is consistent with other studies demonstrating myositis as one of the earliest developing irAEs, with reports of a median onset of 31 days.[15–17] The acute nature of this irAE development is important for providers and patients to be aware of to increase early diagnosis and intervention.

Prior studies have established ICI myositis triad syndrome as a rare and often lethal adverse effect of cancer immunotherapy.[2,4,13,18] Although most patients will develop at least one adverse event secondary to ICI therapy, death from irAEs only represents 0.86% of all treated patients.[14,19] The World Health Organization pharmacovigilance surveys measured the mortality of ICI myositis at 21.2%.[2] ICI triad has mortality ranging from 42% to 51% in retrospective cohort studies.[5,6] Our study further confirms that the prevalence of irAE myositis is low (0.23%), but mortality, specifically with ICI myositis triad, is high (60%). Due to the high mortality, early evaluation for ICI myositis triad should be obtained and supportive care provided.

Patients with ICI triad syndrome were significantly older than patients with isolated ICI myositis, with a median age of 76.1 versus 70.9 years at time of initial ICI infusion. Although rates of low-grade immune toxicities are similar regardless of age, multiple studies have found a higher incidence of common terminology criteria for adverse events grade 3 and higher irAEs in patients 65 years and older.[17,20,21]

Additionally, we found a potential signal for genitourinary cancers for the development of ICI triad syndrome that requires future investigation. Genitourinary cancers were twice as common in ICI triad compared to isolated ICI myositis. Multiple case reports describe patients with ICI triad with genitourinary cancers.[22–25] The patient from the study by Botta et al had elevated activated CD8 T cells and the presence of HLA-DRB1*03/*04 haplotype. If this finding is confirmed in other studies, it could support the theory that ICI triad is secondary to the inappropriate activation of T cells against a shared epitope between tumor cells and healthy muscle.[26] Perhaps urothelial cancer, which is enriched in smooth muscle, increases the chances of immune dysregulation targeting normal myocytes.

Strengths of our study included the availability of the VA standardized national EHR in the largest health care system in the US. We were able to access these EHRs and data collection systems across facilities nationwide. In addition, we were able to compare those with ICI triad syndrome, those with isolated ICI myositis, and those with ICI who did not develop myositis, a novel investigation in this arena.

Limitations of this study include the retrospective study design. Although our definition of ICI myositis was specific, we may have missed cases in which CK was not measured or myositis was not noted in the EHR. Similarly, we may not have captured cases that were labeled as irAE myocarditis or myasthenia gravis, thereby overshadowing a diagnosis of ICI myositis. Veterans who had ICI treatment within VHA may also have developed and been diagnosed with ICI-associated myositis outside VHA and not captured in our analysis. Additionally, our patients were White male veterans, which limits generalizability. Although this report is one the largest reports of ICI-associated myositis to date, the relatively small numbers limit the detection of risk factors. Although efforts were made to adjust for potential confounding factors, we recognized the potential for unidentified residual confounders.

Our study adds to the knowledge of ICI-induced adverse events by describing and comparing ICI triad and isolated ICI myositis subsets in a large national population. ICI triad syndrome is associated with worse overall survival and median survival times

in comparison to patients with isolated ICI myositis and patients treated with ICI without ICI myositis. In contrast, patients with isolated ICI myositis have similar outcomes to patients treated with ICI who did not develop myositis. These data provide useful information to clinicians faced with these irAEs as they communicate expected outcomes. These data also highlight the need for early recognition and treatment of these disease entities.

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 Rubino 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. Shin YE, Kumar A, Guo JJ. Spending, utilization, and price trends for immune checkpoint inhibitors in US Medicaid programs: an empirical analysis from 2011 to 2021. Clin Drug Investig 2023;43(4):289–298. https://doi.org/10.1007/s40261-023-01254-x.
2. Anquetil C, Salem JE, Lebrun-Vignes B, et al. Immune checkpoint inhibitor-associated myositis: expanding the spectrum of cardiac complications of the immunotherapy revolution. Circulation 2018; 138(7):743–745.
3. Nguyễn T, Maria ATJ, Ladhari C, et al. Rheumatic disorders associated with immune checkpoint inhibitors: what about myositis? An analysis of the WHO's adverse drug reactions database. Ann Rheum Dis 2022;81(2):e32.
4. Aldrich J, Pundole X, Tummala S, et al. Inflammatory myositis in cancer patients receiving immune checkpoint inhibitors. Arthritis Rheumatol 2021;73(5):866–874.
5. Shelly S, Triplett JD, Pinto MV, et al. Immune checkpoint inhibitor-associated myopathy: a clinicoseropathologically distinct myopathy. Brain Commun 2020;2(2):fcaa181.
6. Boutros A, Bottini A, Rossi G, et al. Neuromuscular and cardiac adverse events associated with immune checkpoint inhibitors: pooled analysis of individual cases from multiple institutions and literature. ESMO Open 2023;8(1):100791.
7. Hamada N, Maeda A, Takase-Minegishi K, et al; YCU irAE Working Group. Incidence and distinct features of immune checkpoint inhibitor-related myositis from idiopathic inflammatory myositis: a single-center experience with systematic literature review and meta-analysis. Front Immunol 2021;12:803410.
8. Sohn MW, Arnold N, Maynard C, et al. Accuracy and completeness of mortality data in the Department of Veterans Affairs. Popul Health Metr 2006;4:2.
9. Zullig LL, Jackson GL, Dorn RA, et al. Cancer incidence among patients of the U.S. Veterans Affairs Health Care System. Mil Med 2012;177(6):693–701.
10. Bonaca MP, Olenchock BA, Salem JE, et al. Myocarditis in the setting of cancer therapeutics: proposed case definitions for emerging clinical syndromes in cardio-oncology. Circulation 2019;140(2):80–91.
11. Saygin D, Ghosh N, Reid P. Immune checkpoint inhibitor-associated myositis: a distinct form of inflammatory myopathy. J Clin Rheumatol 2022;28(7):367–373.
12. Haratani K, Hayashi H, Chiba Y, et al. Association of immune-related adverse events with nivolumab efficacy in non-small-cell lung cancer. JAMA Oncol 2018;4(3):374–378.
13. Shankar B, Zhang J, Naqash AR, et al. Multisystem immune-related adverse events associated with immune checkpoint inhibitors for treatment of non-small cell lung cancer. JAMA Oncol 2020;6(12): 1952–1956.
14. Tison A, Quéré G, Misery L, et al; Groupe de Cancérologie Cutanée, Groupe Français de Pneumo-Cancérologie, and Club Rhumatismes et Inflammations. Safety and efficacy of immune checkpoint inhibitors in patients with cancer and preexisting autoimmune disease: a nationwide, multicenter cohort study. Arthritis Rheumatol 2019;71(12): 2100–2111.
15. Allenbach Y, Anquetil C, Manouchehri A, et al. Immune checkpoint inhibitor-induced myositis, the earliest and most lethal complication among rheumatic and musculoskeletal toxicities. Autoimmun Rev 2020;19(8):102586.
16. Jayan A, Mammen AL, Suarez-Almazor ME. Immune checkpoint inhibitor-induced myositis. Rheum Dis Clin North Am 2024;50(2): 281–290.
17. Pathak R, Katel A, Massarelli E, et al. Immune checkpoint inhibitor-induced myocarditis with myositis/myasthenia gravis overlap syndrome: a systematic review of cases. Oncologist 2021;26(12):1052–1061.
18. Longinow J, Zmaili M, Skoza W, et al. Immune checkpoint inhibitor induced myocarditis, myasthenia gravis, and myositis: a single-center case series. Cancer Med 2023;12(3):2281–2289.
19. Bertrand A, Kostine M, Barnetche T, et al. Immune related adverse events associated with anti-CTLA-4 antibodies: systematic review and meta-analysis. BMC Med 2015;13(1):211.
20. Elias R, Hartshorn K, Rahma O, et al. Aging, immune senescence, and immunotherapy: a comprehensive review. Semin Oncol 2018; 45(4):187–200.
21. Chen C, Wu B, Zhang C, et al. Immune-related adverse events associated with immune checkpoint inhibitors: an updated comprehensive disproportionality analysis of the FDA adverse event reporting system. Int Immunopharmacol 2021;95:107498.
22. Botta C, Agostino RM, Dattola V, et al. Myositis/myasthenia after pembrolizumab in a bladder cancer patient with an autoimmunity-associated HLA: immune-biological evaluation and case report. Int J Mol Sci 2021;22(12):6246.
23. Todo M, Kaneko G, Shirotake S, et al. Pembrolizumab-induced myasthenia gravis with myositis and presumable myocarditis in a patient with bladder cancer. IJU Case Rep 2019;3(1):17–20.
24. Sessums M, Yarrarapu S, Guru PK, et al. Atezolizumab-induced myositis and myocarditis in a patient with metastatic urothelial carcinoma. BMJ Case Rep 2020;13(12):e236357.
25. Saad R, Ghaddar A, Zeenny RM. Pembrolizumab-induced myocarditis with complete atrioventricular block and concomitant myositis in a metastatic bladder cancer patient: a case report and review of the literature. J Med Case Rep 2024;18(1):107.
26. Chan KK, Bass AR. Autoimmune complications of immunotherapy: pathophysiology and management. BMJ 2020;369:m736.

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1130–1139
DOI 10.1002/acr.70039
© 2025 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

American College of Rheumatology
Empowering Rheumatology Professionals

Mycophenolate Mofetil Treatment Reduces the Risk of Treatment Escalation Due to Vascular Complications in Limited Cutaneous Systemic Sclerosis: Emulation of a Target Trial From the Italian Rheumatology Society SPRING Registry

Enrico De Lorenzis, MD, PhD,[1,2] Gerlando Natalello, MD,[1,2] Rossella De Angelis, MD,[3] Lucrezia Verardi, MD,[1,2] Dilia Giuggioli, MD,[4] Gianluigi Bajocchi, MD,[5] Lorenzo Dagna, MD,[6] Silvia Bellando-Randone, MD,[7] Giovanni Zanframundo, MD,[8,9] Rosario Foti, MD,[10] Fabio Cacciapaglia, MD,[11] Giovanna Cuomo, MD,[12] Alarico Ariani, MD,[13] Edoardo Rosato, MD,[14] Gemma Lepri, MD,[7] Francesco Girelli, MD,[15] Valeria Riccieri, MD,[16] Elisabetta Zanatta, MD,[17,18] Ilaria Cavazzana, MD,[19] Francesca Ingegnoli, MD,[20] Maria De Santis, MD,[21,22] Giuseppe Murdaca, MD,[23] Giuseppina Abignano, MD, PhD,[24] Giorgio Pettiti, MD,[25] Alessandra Della Rossa, MD,[26] Maurizio Caminiti, MD,[27] Annamaria Iuliano, MD,[28] Giovanni Ciano, MD,[29] Lorenzo Beretta, MD,[30] Gianluca Bagnato, MD,[31] Ennio Lubrano, MD,[32] Maria Ilenia De Andres, MD,[33] Alessandro Giollo, MD,[34] Cosimo Bruni, MD, PhD,[7] Martina Orlandi, MD,[5] Marco Fornaro, MD,[35] Marta Saracco, MD,[36] Cecilia Agnes, MD,[37] Pier Giacomo Cerasuolo, MD,[1,2] Gabriella Alonzi, MD,[1,2] Edoardo Cipolletta, MD, PhD,[3] Federica Lumetti, MD,[4] Amelia Spinella, MD,[4] Luca Magnani, MD,[5] Corrado Campochiaro, MD,[6] Giacomo De Luca, MD,[6] Veronica Codullo, MD,[8,9] Elisa Visalli, MD,[10] Carlo Iandoli, MD,[12] Antonietta Gigante, MD,[14] Greta Pellegrino, MD,[38,39] Erika Pigatto, MD,[40] Maria-Grazia Lazzaroni, MD,[19] Franco Franceschini, MD,[19] Elena Generali, MD,[22] Gianna Mennillo, MD,[24] Simone Barsotti, MD,[26] Giuseppa Pagano Mariano, MD,[27] Federica Furini, MD,[41] Licia Vultaggio, MD,[41] Simone Parisi, MD,[42] Clara Lisa Peroni, MD,[42] Gerolamo Bianchi, MD,[43] Enrico Fusaro, MD,[42] Gian Domenico Sebastiani, MD,[28] Marcello Govoni, MD,[41] Salvatore D'Angelo, MD,[24,44] Franco Cozzi, MD,[40] Fabrizio Conti, MD,[16] Serena Guiducci, MD,[7] Andrea Doria, MD,[18] Carlo Salvarani, MD,[45] Florenzo Iannone, MD,[35] Maria Antonietta D'Agostino, MD, PhD,[1,2] Clodoveo Ferri, MD,[4] Marco Matucci Cerinic, MD,[6,46] and Silvia Laura Bosello, MD, PhD,[1,2] on behalf of the SPRING Italian Registry Collaborators

Objective. Mycophenolate mofetil (MMF) use in limited cutaneous systemic sclerosis (lcSSc) is relatively uncommon because of the lower fibrotic burden and the predominance of vascular complications. In vitro observations and clinical data from transplanted patients suggest a protective effect of MMF on endothelial function. Our aim was to evaluate the reasons for prescribing MMF treatment in patients with lcSSc and its impact on the need for escalation of vascular complication–related treatments during follow-up.

Methods. Patients with lcSSc enrolled in the Italian Systemic Sclerosis Progression Investigation registry were retrospectively evaluated. All patients treated with MMF were matched to patients not treated with MMF, which was based on a roll-entry time-dependent propensity score built on demographics, clinical features, and baseline treatment. The escalation of vasoactive or vasodilator treatment up to 60 months was defined as the introduction of iloprost, endothelin receptor antagonists, or phosphodiesterase-5 inhibitors on top of the ongoing treatment, because of uncontrolled or newly diagnosed vascular complications. A hazards Cox model was also adopted to quantify the association of MMF treatment with treatment escalation.

Results. A total of 1,435 patients with lcSSc were evaluated, of whom 152 were prescribed MMF (17.1% male; mean age at lcSSc onset 48.7 ± 13.9 years, 54.6% anti-Scl70 positive). The prescription of MMF was more common in men and in anti-Scl70 positive, anticentromere negative patients with interstitial lung disease, myositis, and without a history of digital ulcers. After matching 107 patients with MMF-untreated controls, the overall incidence of

vasoactive/vasodilator treatment escalation events related to digital ulcers over a median follow-up of 40.5 months (interquartile range 23.3–60.0) was 0.3 per 100 patient-years in the MMF-treated group and 5.4 per 100 patient-years in the matched control group, with a significant difference in treatment escalation-free survival between the two groups (hazard ratio 0.05, 95% confidence interval 0.01–0.38; *P* value = 0.004).

Conclusion. In patients with lcSSc, the introduction of MMF has reduced the need for escalation of vasoactive or vasodilator treatment, suggesting that it may also help to prevent vascular complications, which frequently affect patients with lcSSc.

INTRODUCTION

Systemic sclerosis (SSc) is a chronic autoimmune disease characterized by vascular, immune, and fibrotic complications that can affect the skin, the musculoskeletal system, and the internal organs. The limited cutaneous SSc (lcSSc) subset is the most frequent disease form[1] and accounts for three of four patients. It features a reduced skin extension and a less frequent or severe presentation of pulmonary and cardiac fibrosis. LcSSc, in particular, is associated with significant vascular manifestations such as Raynaud phenomenon, digital ulcers (DUs), and pulmonary arterial hypertension (PAH).[2,3] The sine-scleroderma subset is close to the lcSSc subset because of the absence of skin involvement but it has similar vascular and visceral patterns that contribute to defining SSc as a vascular disease characterized by a unified vascular phenotype.[4,5]

Based on available randomized controlled trials,[6,7] immunosuppression is only recommended for patients with SSc with established inflammatory and fibrotic complications because it demonstrates a better outcome in this subset as compared with placebo. As a result, patients with lcSSc are less frequently treated with these medications as compared with diffuse cutaneous SSc (dcSSc), for which these complications are far more common. This therapeutic approach seems counterintuitive, given that immune activation in SSc is extensively connected to microvascular impairment and fibrosis[8] across disease subsets.

Among the available immunosuppressants, mycophenolate mofetil (MMF) is extensively used as first-line treatment for SSc interstitial lung disease (ILD), and it remains an option for severe cutaneous involvement or SSc-associated myositis. In animal models and in vitro, MMF has been shown to control some aspects of the pathogenesis of SSc vascular complications,

[1]Division of Rheumatology and Clinical Immunology, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy; [2]Rheumatology, Faculty of Medicine and Surgery, Catholic University of the Sacred Heart, Rome, Italy; [3]Rheumatology Unit, Department of Clinical and Molecular Sciences, Polytechnic University of Marche, Ancona, Italy; [4]Department of Medical and Surgical Sciences for Children and Adults, University Hospital of Modena and Reggio Emilia School of Medicine, Modena, Italy; [5]Rheumatology Unit, Azienda USL IRCCS di Reggio Emilia, Reggio Emilia, Italy; [6]Unit of Immunology, Rheumatology, Allergy and Rare Diseases, IRCCS San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy; [7]Division of Rheumatology, Scleroderma Unit, Department of Experimental and Clinical Medicine, AOU Careggi, University of Florence, Florence, Italy; [8]Department of Internal Medicine and Therapeutics, Università di Pavia, Pavia, Italy; [9]Division of Rheumatology, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy; [10]Rheumatology Unit, AOU Policlinico S. Marco, Catania, Italy; [11]Rheumatology Service "F. Miulli" General Hospital Acquaviva delle Fonti - Department of Medicine and Surgery, LUM University "Giuseppe De Gennaro" Casamassima, Bari, Italy; [12]Department of Precision Medicine, University of Campania Luigi Vanvitelli, Caserta, Italy; [13]Department of Medicine, Internal Medicine and Rheumatology, Azienda Ospedaliero Universitaria di Parma, Parma, Italy; [14]Department of Translational and Precision Medicine, Sapienza University of Rome, Roma, Italy; [15]Rheumatology Unit, AUSL Romagna, Ospedale G.B. Morgagni, Forlì, Italy; [16]Department of Rheumatology, Sapienza University of Rome, Rome, Italy; [17]Department of Rheumatology, University of Padua, Padua, Italy; [18]Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padua, Padua, Italy; [19]Rheumatology and Clinical Immunology, ASST Spedali Civili of Brescia; Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy; [20]Rheumatology Clinic, ASST Pini CTO, Department of Clinical Sciences & Community Health, Università degli Studi di Milano, Milan, Italy; [21]Rheumatology and Clinical Immunology, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy; [22]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy; [23]Research Hospital San Martino, University of Genoa, Genoa, Italy; [24]Rheumatology Institute of Lucania (IReL) and Rheumatology Department of Lucania, San Carlo Hospital, Potenza, Italy; [25]Azienda Ospedaliera S Croce e Carle, Cuneo, Italy; [26]Department of Rheumatology, University of Pisa, Pisa, Italy; [27]Departmental Rheumatology Unit, Grande Ospedale Metropolitano, Reggio Calabria, Italy; [28]Rheumatology Unit, San Camillo - Forlanini Hospital, Rome, Italy; [29]Hospital of Ariano Irpino, Local Health Department, Ariano Irpino, Avellino, Italy; [30]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano, Milan, Italy; [31]Department of Clinical and Experimental Medicine, University of Messina, Messina, Italy; [32]Department of Rheumatology, University of Molise, Campobasso, Italy; [33]Rheumatology Unit, Azienda Ospedaliera di Rilievo Nazionale e di Alta Specializzazione "Garibaldi," Catania, Italy; [34]Rheumatology Section, Department of Medicine, University of Verona, Verona, Italy; [35]Rheumatology Unit, Department of Precision and Regenerative Medicine Ionian Area, University of Bari Aldo Moro, Bari, Italy; [36]Rheumatology Unit, Mauriziano-Umberto I Hospital, Torino, Italy; [37]San Lorenzo Hospital, Carmagnola, Turin, Italy; [38]IRCCS Ospedale Galeazzi Sant'Ambrogio, Milan, Italy; [39]Dipartimento di Scienze Biomediche e Cliniche, Università degli Studi di Milano, Milan, Italy; [40]Medicine Unit, Department of Medicine, San Bassiano Hospital, Bassano del Grappa, Italy; [41]Rheumatology Unit, Department of Medical Sciences, University of Ferrara and Azienda Ospedaliera-Universitaria S. Anna di Ferrara, Ferrara, Italy; [42]Rheumatology Unit, University Hospital Città della Salute e della Scienza di Torino, Turin, Italy; [43]Rheumatology Unit, Department of Musculoskeletal Sciences, Local Health Trust La Colletta Hospital, Genoa, Italy; [44]Department of Health Science, University of Basilicata, Potenza, Italy; [45]Rheumatology Unit, Azienda USL IRCCS di Reggio Emilia, University of Modena and Reggio Emilia, Reggio Emilia, Italy; [46]Inflammation Fibrosis and Ageing Initiative (INFLAGE), Division of Genetics and Cell Biology, IRCCS San Raffaele Scientific Institute, Milan, Italy.

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.70039).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.70039.

Address correspondence via email to Gerlando Natalello, MD, at gerlando.natalello@guest.policlinicogemelli.it.

Submitted for publication May 2, 2025; accepted in revised form November 5, 2025.

including endothelial activation and vascular intimal proliferation.[9–13] MMF has also been suggested to exert benefits on the microcirculation of transplanted patients when compared with alternative immunosuppressive regimens.[8,14,15] Our aim was to investigate the reason for introducing MMF treatment and its impact on the escalation of vasoactive or vasodilatory treatment due to vascular complications in patients with a lcSSc phenotype who were enrolled in a large national cohort in the Italian Society of Rheumatology (SIR)–Systemic Sclerosis Progression INvestiGation (SPRING) registry. This study represents the first large, propensity score–matched analysis to explore a rationale for MMF impact on vascular outcomes in lcSSc. The results demonstrate a marked reduction in the need for escalation of vasoactive or vasodilator therapy among MMF-treated patients, indicating a possible vascular-protective role of MMF beyond its established activity.

METHODS

Study design and participants. The research design and the results report adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) initiative guidelines.[16] The study involved a quasi-experimental cohort comparison of patients exposed to MMF alongside a parallel control group not exposed to MMF, which were matched using a roll-entry strategy based on a dynamic, time-dependent propensity score.

The eligible patients were enrolled in the SPRING registry, a multicenter, nonprofit national cohort study initiated by the SIR. This registry, which focused on patients with SSc, involved a total of 38 national tertiary centers with specific expertise in managing the condition.[17] The study protocol was approved by the local ethical committees of each participating center, and all patients provided written informed consent.

All patients included in the analysis met the American College of Rheumatology/EULAR 2013 classification criteria for SSc,[18] were classified as lcSSc according to LeRoy et al criteria,[1] and had available information regarding MMF exposure. Electronic health records were evaluated using a standard workflow specifically developed for the SPRING research protocol and fulfilled by the clinicians directly in charge of the patients with SSc.[19]

Data collection was conducted using REDCap electronic tools. Variables of interest included demographics, the Charlson Comorbidity Index (CCI) at the end of observation,[20] age at SSc onset, disease duration, presence of anticentromere antibodies (ACA) and anti-Scl70 antibodies, and disease domains involved. These domains included ILD, which was defined as parenchymal involvement over 10% on high-resolution computed tomography (HRCT), presence of DUs, myositis, and cardiac involvement. The severity of lung involvement was assessed based on the most recent forced vital capacity (FVC) and the diffusing capacity of the lungs for carbon monoxide (DL_{CO}). For each previous and ongoing treatment, the start and end dates were recorded in the registry.

Rationale of exposure and endpoint definitions. For the quasi-experimental analysis, we designated three months as the minimum treatment duration for a patient to be considered exposed to MMF, considering the common practice of dose titration and the anticipated delay between initiating this medication and observing any clinical benefits for labeled MMF use. In the same analysis, we chose the escalation of vasodilator or vasoactive treatment as the endpoint. This escalation was defined as the introduction of intravenous iloprost, endothelin receptor antagonists, or phosphodiesterase type 5 inhibitors (PDE5i) to the existing treatment regimen due to uncontrolled or newly diagnosed vascular manifestations (eg, severe Raynaud phenomenon, DUs or pitting scars, and PAH). The initiation of calcium channel blockers (CCB) was not categorized as an event because of its established role as standard background therapy for most patients with SSc, irrespective of the development of vascular complications. Nevertheless, CCB was regarded as a potential confounder in statistical analysis.

The preference for therapeutic escalation as an endpoint over the formal diagnosis of new vascular complications was based on its expected robustness to minimize selection, information, recall, observer, and misclassification biases. This approach acknowledges the difficulties of retrospectively defining a worsening of acral vascular disease in a reliable way and the challenges in accessing right heart catheterization in real-world settings.

Statistical analysis. Categorical variables were presented as numbers and percentages, whereas continuous variables were reported as mean ± SD or median with interquartile range (IQR), depending on the normality of the data, which were assessed through the inspection of quantile-quantile plots. Cross-sectional comparisons between patient groups receiving MMF and those not receiving MMF were conducted using the chi-square test or Fisher exact test for categorical variables and the Mann-Whitney U test or *t*-test for continuous variables, as appropriate.

A roll-entry strategy was adopted to select a matched control group.[21] The duration of disease from diagnosis to the last follow-up was divided into four-month intervals for each patient. We selected four-month intervals because this period reflects the average frequency of follow-up assessments in clinical practice for patients with SSc, making it a clinically relevant timeframe for monitoring disease evolution. Moreover, this interval was expected to balance sufficient granularity for detecting clinically meaningful changes while avoiding excessive fragmentation of the data. A dynamic propensity score was calculated for each four-month period using a binomial logistic model that incorporated both time-independent and time-dependent variables evaluated at the end of the previous four-month period. Time-

dependent variables included age, disease duration from the first symptoms (excluding Raynaud phenomenon), and baseline treatment with CCBs, bosentan, macitentan, ambrisentan, sildenafil, tadalafil, rituximab, and tocilizumab, which were individually considered. Time-independent variables included sex, anti-Scl70 positivity, ACA positivity, and CCI. The last available FVC and DL_{CO} values were also approximated as time-independent matching variables that served as general proxies for overall comorbidity burden and severity of lung involvement. We evaluated patients who were MMF-exposed in the four-month period of medication initiation and finally matched each of them with control patients who were among those evaluated in the same four-month period along the timeline. Some patients were preliminarily excluded to prevent variable imbalance if their propensity scores differed significantly from the pooled propensity score of the treatment group. Patients were excluded if their propensity scores exceeded a prespecified fraction of the pooled propensity score SD that was defined by a caliper coefficient set at one SD of the pooled propensity score of the MMF-treated group. The matching was one-to-one without replacement and based on the closest dynamic propensity score for that specific four-month period (Figure 1). Standardized mean differences (SMDs) after matching for variables in the propensity score model were reported as a measure of balance between the intervention and comparison group with values 0.1 or below indicating a good balance.[22]

Kaplan-Meier survival analysis was conducted to compare patients who were MMF-exposed and the matched control group in preventing the endpoint based on the Log-Rank test. Censoring rate and pattern similarity assumptions were assessed. A

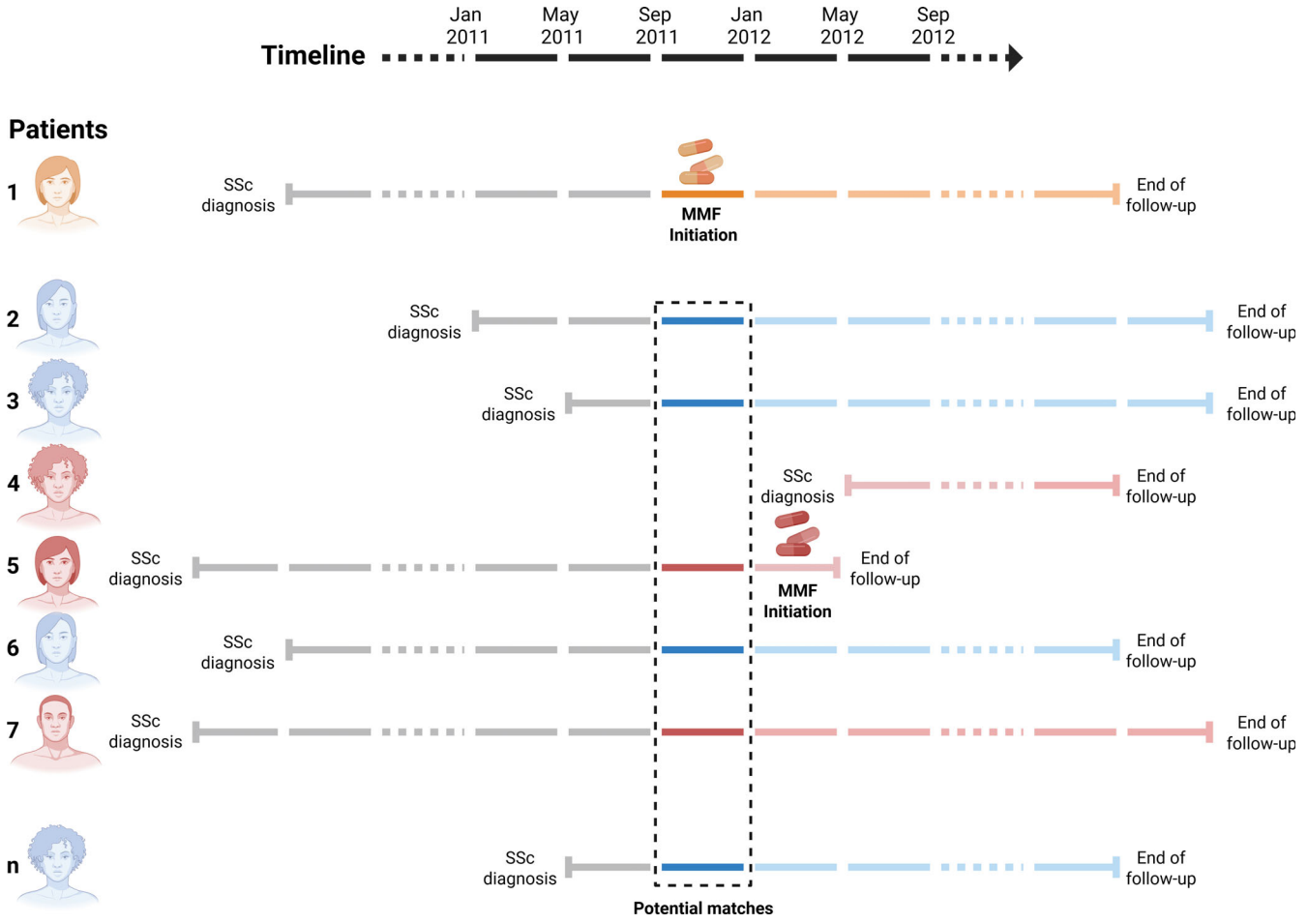

Figure 1. Example roll-entry matching strategy: Patient 1 initiated MMF treatment in the four-month period spanning September 2011 to January 2012. Patient 5, along with any other patient treated with MMF for at least three months, could not be matched. Similarly, patient 4, and any other patient diagnosed after that four-month period or concluding the follow-up before that four-month period, could not be matched. The propensity score of patient 1 calculated in the same four-month period was compared with that of the remaining patients. Patient 7, along with any other patients whose propensity score is too distant from that of patient 1 according to the established caliper, was excluded a priori. Finally, patient 1 was matched with patients among the remaining group with the closest propensity score calculated for that four-month period. Color code: orange (MMF-treated patients to be matched), blue (MMF-untreated patients selected for matching), red (patients who could not be matched). Gray segments indicate disease intervals preceding the four-month period of interest. Created with BioRender.com. MMF, mycophenolate mofetil; SSc, systemic sclerosis.

univariate proportional hazards Cox model was also adopted to quantify the association of MMF treatment and the endpoint given group balance of clinical variables. Results were presented as hazard ratios (HR) with 95% confidence intervals (CI). The proportional hazards assumption was verified by excluding statistically significant correlations between scaled Schoenfeld residuals and time. For both analyses, the observation period ended at the occurrence of an event, discontinuation of MMF for the treatment group, loss to follow-up, or at the 60th month.

Statistical significance was defined as a *P* value less than 0.05 for all analyses, and all tests were two-tailed. A Bonferroni correction was performed in case of multiple comparisons. Data analysis was performed using RStudio.

Bias adjustment and missing data handling. Multiple bias mitigation strategies were adopted to enhance validity and reliability. First, uniform methods and tools for retrospective data collection across all study centers and patients were implemented to mitigate information bias according to the SPRING research protocol. Second, for the preliminary cross-sectional comparison of patients receiving or not receiving MMF, a sensitivity analysis was conducted for patient subgroups diagnosed in three time periods: up to the publication of the 2009 EULAR systemic sclerosis guidelines,[23] after the 2017 guidelines update,[24] and between these two time periods. This approach was aimed at minimizing temporal and selection biases by considering the evolution of the standard of care over time and by retrospectively considering patient subgroups with potentially different survival outcomes. Third, the use of dynamic time-dependent propensity score matching in the definition of longitudinal analysis was aimed to dampen temporal biases because compared patients were considered within the same four-month period of the timeline. Matching variables included both time-dependent and time-independent variables.

Patients with missing information on vasoactive or vasodilator treatment were excluded from the longitudinal analysis because they were directly involved in the endpoint assessment. Predictive mean matching was used as an imputation method for missed FVC and DLco based on remaining collected predictors and the generation of five imputed data sets. An additional sensitivity analysis based on only completed observations was proposed.

Minimal sample size definition. It was determined that a minimum of 60 patients in each group would be required to demonstrate a 50% risk reduction (HR = 0.5) in vasoactive or vasodilator treatment escalation with MMF treatment compared with a control group of the same size. This calculation assumed a two-tailed alpha of 0.05 and a power of 0.9 based on an expected baseline event rate of two events per year in the control group[25] and an average follow-up duration of 2.5 years. Additionally, it accounted for an annual censoring rate of 50% in both the intervention and control groups.[26]

Table 1. Comparison of clinical variables between patients with lcSSc who were and were not receiving MMF at the last available follow-up*

	All patients	Patients not receiving MMF	Patients receiving MMF	*P* value
N	1,435	1,283	152	–
Men, n (%)	133 (9.3)	107 (8.3)	26 (17.1)	**<0.001**
Last available CCI, mean ± SD	2.4 ± 1.7	2.4 ± 1.7	2.2 ± 1.5	0.2
Age at SSc onset, mean ± SD, y	49.8 ± 14.2	49.9 ± 14.2	48.7 ± 13.9	0.3
ACA positivity, n (%)	604 (42.1)	589 (45.9)	15 (9.9)	**<0.001**
Anti-Scl70 positivity, n (%)	395 (27.5)	312 (24.3)	83 (54.6)	**<0.001**
ILD on HRCT, n (%)	573 (39.9)	449 (35.0)	124 (81.6)	**<0.001**
Last available FVC, % of predicted, mean ± SD	103.8 ± 21.8	105.6 ± 20.7	88.8 ± 24.8	**<0.001**
Last available DLco, % of predicted, mean ± SD	69.0 ± 19.4	70.0 ± 18.9	60.8 ± 21.7	**<0.001**
DUs (ever), n (%)	327 (22.8)	303 (23.6)	24 (15.8)	0.030
Myositis, (%)	15 (1.0)	10 (0.8)	5 (3.3)	**0.016**
Baseline CCB, n (%)	961 (67.0)	857 (66.8)	104 (68.4)	0.7
Baseline sildenafil, n (%)	49 (3.4)	42 (3.3)	7 (4.6)	0.4
Baseline tadalafil, n (%)	16 (1.1)	15 (1.2)	1 (0.7)	>0.9
Baseline bosentan, n (%)	329 (22.9)	287 (22.4)	42 (27.6)	0.14
Baseline macitentan, n (%)	32 (2.2)	28 (2.2)	4 (2.6)	0.8
Baseline ambrisentan, n (%)	9 (0.6)	9 (0.7)	0 (0.0)	0.6
Baseline iloprost, n (%)	718 (50.0)	646 (50.4)	72 (47.4)	0.5
Baseline rituximab, n (%)	33 (2.3)	25 (1.9)	8 (5.3)	**0.018**
Baseline tocilizumab, n (%)	28 (2.0)	25 (1.9)	3 (2.0)	**>0.9**

* Bold *P* values indicate statistically significant differences with the threshold set at 0.0025 after Bonferroni adjustment for multiple comparisons. ACA, anti-centromere antibody; CCB, calcium channel blocker; CCI, Charlson Comorbidity Index; DU, digital ulcer; DLco, diffusing capacity of the lung for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution computed tomography; ILD, interstitial lung disease; IQR, interquartile range; lcSSc, limited cutaneous systemic sclerosis; MMF, mycophenolate mofetil.

The study was conducted within the ethical approval from each participating center according to the SPRING Italian registry policy. The authors confirm that the data supporting the findings of this study are available within the article or its supplementary materials. Patients were not directly involved in setting the research question, the endpoint measures, or the design of this study. However, unmet clinical needs that emerged during routine clinical practice were considered. Patients were informed of the results of this study on request.

RESULTS

Patients' characteristics and clinical comparisons of patients receiving and not receiving MMF. The patient selection process is summarized in Supplementary Figure 1. A total of 1,435 patients with lcSSc were enrolled in the SPRING registry as of December 31, 2022. The patients resided throughout the country, particularly in proximity to the main urban centers, where the enrollment sites were located (Supplementary Figure 2A).

The median observation time from the diagnosis of SSc to the last available follow-up was 105 months (IQR 54–180). MMF was prescribed to 152 (10.6%) patients within the timeframe from diagnosis to the last follow-up, with a median time from diagnosis to MMF prescription of 28 (IQR 5–82) months. The timeline of MMF prescription is reported in Supplementary Figure 2B.

The baseline clinical characteristics of the overall cohort and the comparison of patients receiving and not receiving MMF before matching are summarized in Table 1. After adjusting for multiple comparisons, patients receiving MMF were more commonly men and presented a higher frequency of a history of ILD and myositis, as well as being anti-Scl70 positive and ACA negative. The statistical association between MMF prescription with no history of DUs was lost when adjusted for multiple comparisons, although patients with DUs tended to be less likely to have been exposed to MMF compared with patients without any history of these complications due to concomitant organ involvements. Patients who were receiving MMF were also more likely to have been exposed to rituximab. Subgroup analysis confirmed these associations when patients diagnosed before 2007, from 2008

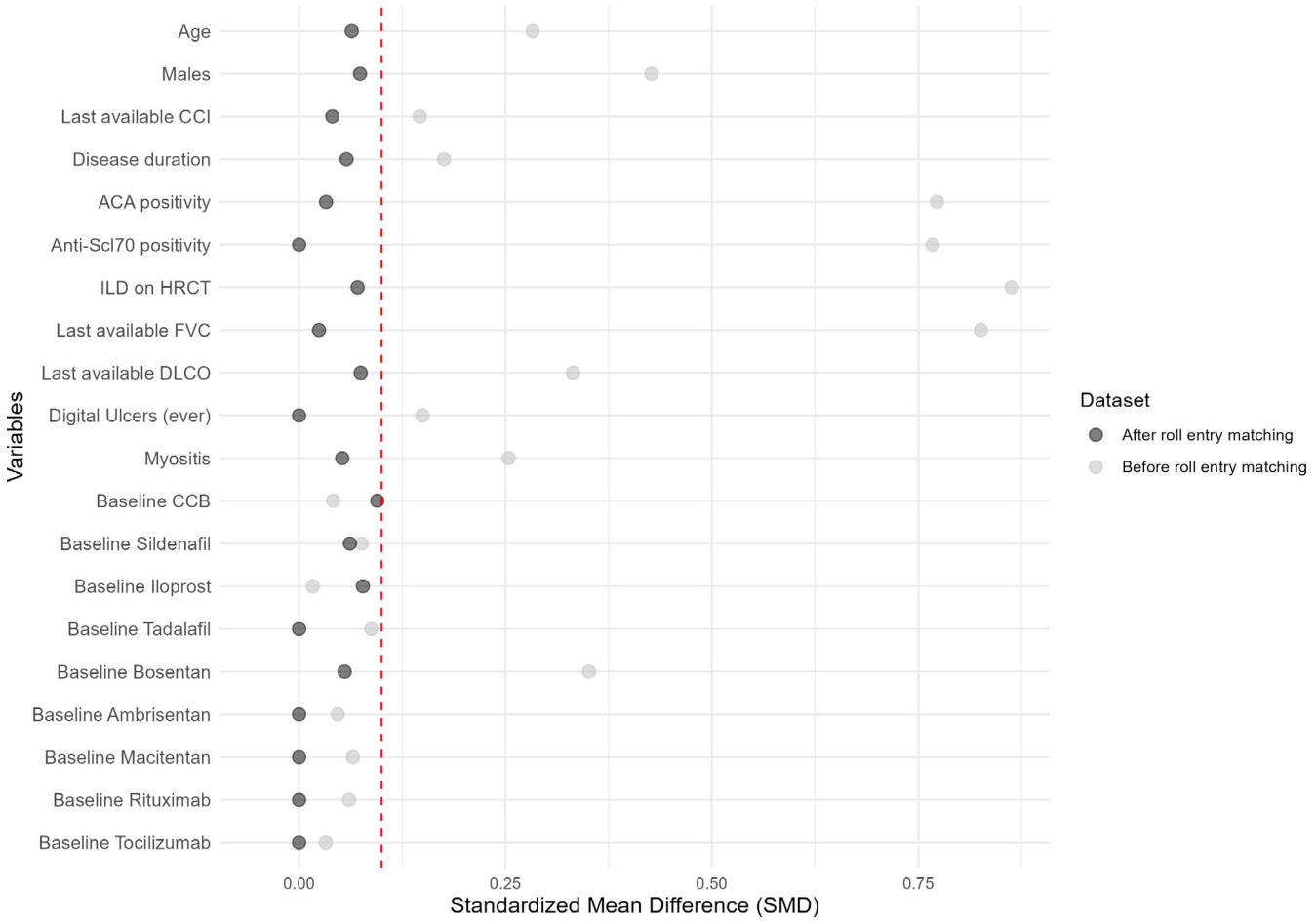

Figure 2. Love plot showing balancing covariates before and after roll-entry matching. ACA, anti-centromere antibody; CCB, calcium channel blocker; CCI, Charlson Comorbidity Index; DLco, diffusing capacity of the lung for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution computed tomography; ILD, interstitial lung disease. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.70039/abstract.

to 2015, and after 2016 were considered in the subgroup analysis, as shown in Supplementary Tables 1 to 3.

MMF treatment and vasoactive or vasodilator treatment escalation. Out of 152 patients who received MMF, 19 discontinued the medication within the first three months, and 26 were excluded because of incomplete information in follow-up. Consequently, 107 patients remained and were matched with 107 controls who were unexposed to MMF using a time-dependent propensity score. This matching was deemed successful based on SMD values under 0.1, as summarized in Figure 2 and detailed in Table 2. In a quasi-experimental design, a total of 214 patients were followed for a median duration of 40.5 months (IQR 23.3–60.0) after the index date, which was the date of MMF initiation in the intervention group and the date of matching for the control group.

During follow-up, 18 instances of escalated vasoactive or vasodilator treatment were recorded. This corresponds to an incidence rate of 2.2 events per 1,000 patient-years. Escalation of treatment was due to uncontrolled DUs in 15 patients and severe Raynaud phenomenon in 3 patients. Treatment initiation included iloprost initiation in 11 patients, a combination of bosentan and iloprost initiation in 6 patients, and bosentan initiation in 1 patient (Supplementary Table 4). One single event was reported in the MMF-treated group (0.3 per 100 patient-years), whereas 17 occurred in the control group (5.4 per 100 patient-years). At the end of follow-up, the cumulative incidence of events was 2.5% (95% CI 0.0–7.2) in the MMF group versus 21.8% (95% CI 10.0–32.1) in the control group. The survival distributions between the two groups were significantly different, as indicated by the log-rank test (*P* value = 0.00025) (Figure 3). Univariate Cox regression analysis revealed that MMF treatment was associated with a 95% reduction in risk (HR 0.05, 95% CI 0.01–0.38; *P* value = 0.004) of vascular complications requiring escalation of vasoactive/vasodilator therapy when compared with the matched control group with comparable clinical characteristics.

In the sensitivity analysis, which included only 86 patients with complete clinical observations for each group, the estimated cumulative incidence of vasoactive or vasodilator treatment was 1.2% (IQR 0.0%–7.3%) in the MMF group and 22.8% (10.0%–33.1%) in the control group. A statistically significant difference was confirmed (*P* value <0.001), corresponding to a HR of 0.06 (95% CI 0.01–0.47; *P* value = 0.007) in the Cox analysis (Supplementary Figure 3).

DISCUSSION

Our results show that in the last decades patients with lcSSc were more likely to receive MMF, although in our retrospective cohort only 1 of 10 received MMF. Most importantly, our data suggest that MMF exerts a protective effect against vascular complications. In our cohort, the main clinical characteristic associated with MMF treatment appeared to be ILD, followed by the occurrence of myositis. Consistently, ILD-related risk factors,

Table 2. Comparison of clinical variables between patients with lcSSc who were and were not prescribed MMF during follow-up after roll-entry matching*

	Patients not prescribed MMF	Patients prescribed MMF	SMD
N	107	107	–
Men, n (%)	20 (18.7)	17 (15.9)	0.074
Age, mean ± SD	53.78 ± 14.18	54.64 ± 12.74	0.064
Last available CCI, mean ± SD	2.31 ± 1.64	2.24 ± 1.61	0.040
Disease duration, mean ± SD	5.49 ± 5.01	5.83 ± 6.88	0.057
ACA positivity, n (%)	10 (9.3)	9 (8.4)	0.033
Anti-Scl70 positivity, n (%)	65 (60.7)	65 (60.7)	<0.001
ILD on HRCT, n (%)	85 (79.4)	88 (82.2)	0.071
Last available FVC, % of predicted, mean ± SD	85.57 ± 18.95	89.03 ± 25.49	0.024
Last available DLco, % of predicted, mean ± SD	62.14 ± 17.70	60.64 ± 22.12	0.075
DUs (ever), n (%)	21 (19.6)	21 (19.6)	<0.001
Myositis, n (%)	4 (3.7)	3 (2.8)	0.053
Baseline CCB, n (%)	46 (43.0)	41 (38.3)	0.095
Baseline sildenafil, n (%)	3 (2.8)	2 (1.9)	0.062
Baseline tadalafil, n (%)	0 (0.0)	0 (0.0)	<0.001
Baseline bosentan, n (%)	15 (14.0)	13 (12.1)	0.055
Baseline ambrisentan, n (%)	0 (0.0)	0 (0.0)	<0.001
Baseline macitentan, n (%)	0 (0.0)	0 (0.0)	<0.001
Baseline iloprost, n (%)	41 (38.3)	37 (34.6)	0.078
Baseline rituximab, n (%)	0 (0.0)	0 (0.0)	<0.001
Baseline tocilizumab, n (%)	0 (0.0)	0 (0.0)	<0.001

* ACA, anti-centromere antibody; CCB, calcium channel blocker; CCI, Charlson Comorbidity Index; DU, digital ulcer; DLco, diffusing capacity of the lung for carbon monoxide; FVC, forced vital capacity; HRCT, high-resolution computed tomography; ILD, interstitial lung disease; IQR, interquartile range; lcSSc, limited cutaneous systemic sclerosis; MMF, mycophenolate mofetil; SMD, standardized mean difference.

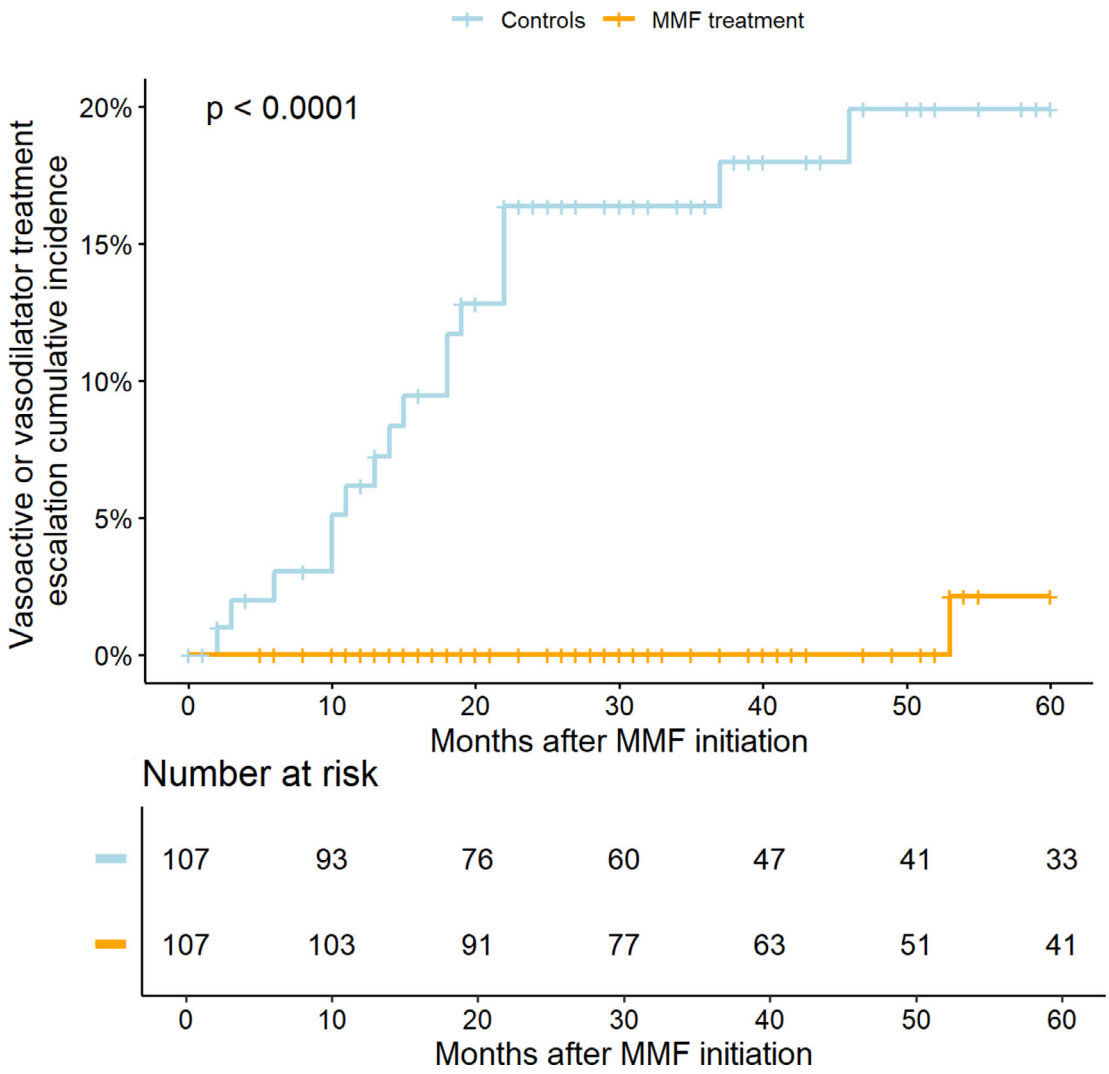

Figure 3. Comparison of cumulative incidence curves of vasoactive or vasodilator treatment escalation in MMF-treated patients and matched controls. MMF, mycophenolate mofetil.

such as male sex, anti-Scl70 positivity, and ACA negativity, were more frequently observed in the MMF receiving group. Interestingly, neither comorbidity burden nor age seemed to influence the prescription of MMF. Additionally, we investigated whether there was any added benefit in preventing the need for a vascular or vasoactive treatment upgrade, wich served as a surrogate event for the onset of uncontrolled vascular disease. Using a propensity score matching strategy, which is a commonly used strategy to simulate randomization in observational studies, we defined a control group. Traditionally, propensity score matching designs are cross-sectional, matching on covariates before the intervention and measuring endpoints after the intervention to analyze the effect of treatment at a specific point in time. Although effective in many situations, this approach assumes that covariates do not change in a relevant time window, or if they do, that these changes will not affect the endpoint variable. Additionally, identifying the index date for the unexposed control group can introduce substantial biases. Time-dependent propensity score matching, associated with a rolling entry strategy, overcomes these limitations and moves closer to a classical randomized study design.

The results of this comparison, corroborated by the sensitivity analysis, indicate that MMF significantly decreases the risk of upgrading to vasoactive or vasodilator treatment. The recorded events were attributed to acral vascular complications. This is not surprising because MMF tended to be initiated early after diagnosis, resulting in shorter disease duration among control group patients, as well. Moreover, because the matching strategy favored patients with similar characteristics, those on macitentan, ambrisentan, tadalafil, and sildenafil, representing a minority group with PAH, were more difficult to match and therefore were less represented in the selection process.

This finding is fully consistent with both preclinical studies reporting the protective effect of MMF on endothelial activation

and intimal proliferation in both in vitro and animal models,[5–9] as well as with the beneficial effect reported in patients who received MMF to prevent solid organ rejection.[8,10,11] Moreover, these data support a more unified vision of SSc pathophysiology in which, other than vasoactive and vasodilators therapy, dampening the immune response could counteract microvascular damage underlying SSc.

The MINIMISE randomized controlled trial is currently evaluating the effects of MMF in a lcSSc population with no clinical recommended indication for an immunosuppressive treatment (exclusion criteria are the presence of lung or cardiac involvement) (EudraCT: 2019-004139-21). It will be interesting to see if in this population MMF will also have protective effect on the vascular compartment because in our study the patients with lcSSc who underwent MMF were those with the presence of organ involvement (ie, characterized by a significant pro-inflammatory and profibrotic burden). Therefore, we cannot exclude the possibility that the effect of MMF on the microvascular compartment was not direct, but rather mediated through the control of fibrotic and immune burden.

Some limitations should be acknowledged. First, propensity score matching cannot account for unmeasured or unknown confounding variables, which potentially bias the estimated treatment effect. Second, the retrospective design may have affected the representativeness of the population examined. Specifically, only long-term survivors diagnosed earlier in the evaluated timeline were included, which may limit the generalizability of the findings. Finally, smoke exposure and the specific MMF dose were not available. Similarly, the specific predominant reason for MMF introduction (eg, ILD, myositis, myocarditis, and early disease with risk factors for the development of a diffuse form) could not be retrospectively reconstructed.

In conclusion, our retrospective analysis indicated that, during the study period, MMF treatment in lcSSc was limited to patients with risk factors and visceral or musculoskeletal involvement approaching the severity characteristic of dcSSc. In our study, MMF treatment was associated with a reduced need for vasoactive or vasodilative therapy escalation, suggesting a beneficial effect on SSc microvascular complications. Randomized controlled studies are warranted to confirm the effects of MMF on vascular SSc manifestations.

ACKNOWLEDGMENTS

We thank the SPRING Italian registry collaborators for their contributions to data collection and the Italian Society of Rheumatology for supporting 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 Natalello 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. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988; 15(2):202–205.
2. Matucci-Cerinic M, Kahaleh B, Wigley FM. Review: evidence that systemic sclerosis is a vascular disease. Arthritis Rheum 2013;65(8): 1953–1962. doi:10.1002/art.37988
3. Denton CP, Khanna D. Systemic sclerosis. Lancet 2017;390(10103): 1685–1699. doi:10.1016/S0140-6736(17)30933-9
4. Allanore Y, Distler O, Matucci-Cerinic M, et al. Review: defining a unified vascular phenotype in systemic sclerosis. Arthritis Rheumatol 2018;70(2):162–170. doi:10.1002/art.40377
5. De Angelis R, Ferri C, Giuggioli D, et al; SPRING-SIR (Systemic Sclerosis PRogression INvestiGation group of the Italian Society of Rheumatology). Systemic sclerosis sine scleroderma: clinical and serological features and relationship with other cutaneous subsets in a large series of patients from the national registry 'SPRING' of the Italian Society for Rheumatology. RMD Open 2023;9(1):e002890. doi:10.1136/rmdopen-2022-002890
6. Distler O, Highland KB, Gahlemann M, et al; SENSCIS Trial Investigators. Nintedanib for systemic sclerosis-associated interstitial lung disease. N Engl J Med 2019;380(26):2518–2528. doi:10.1056/NEJMoa1903076
7. Tashkin DP, Roth MD, Clements PJ, et al; Sclerodema 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. doi:10.1016/S2213-2600(16)30152-7
8. Truchetet ME, Brembilla NC, Chizzolini C. Current Concepts on the Pathogenesis of Systemic Sclerosis. Clin Rev Allergy Immunol 2023; 64(3):262–283. doi:10.1007/s12016-021-08889-8
9. Haug C, Schmid-Kotsas A, Linder T, et al. Influence of hepatocyte growth factor, epidermal growth factor, and mycophenolic acid on endothelin-1 synthesis in human endothelial cells. Nephrol Dial Transplant 2001;16(12):2310–2316. doi:10.1093/ndt/16.12.2310
10. Shimizu H, Takahashi M, Takeda S, et al. Mycophenolate mofetil prevents transplant arteriosclerosis by direct inhibition of vascular smooth muscle cell proliferation. Transplantation 2004;77(11):1661–1667. doi:10.1097/01.TP.0000127592.13707.B6
11. Huang Y, Liu Z, Huang H, et al. Effects of mycophenolic acid on endothelial cells. Int Immunopharmacol 2005;5(6):1029–1039. doi:10.1016/j.intimp.2005.01.015
12. Krötz F, Keller M, Derflinger S, et al. Mycophenolate acid inhibits endothelial NAD(P)H oxidase activity and superoxide formation by a Rac1-dependent mechanism. Hypertension 2007;49(1):201–208. doi:10.1161/01.HYP.0000251162.14782.d4
13. Fréguin-Bouilland C, Godin M, Bellien J, et al. Protective effect of mycophenolate mofetil on endothelial function in an aortic allograft model. Transplantation 2011;91(1):35–41. doi:10.1097/TP.0b013e3181fe12d6
14. Weis M, Wildhirt SM, Schulze C, et al. Coronary vasomotor dysfunction in the cardiac allograft: impact of different immunosuppressive regimens. J Cardiovasc Pharmacol 2000;36(6):776–784. doi:10.1097/00005344-200012000-00014

15. Ruben S, Kreuzer M, Büscher A, et al. Impaired microcirculation in children after kidney transplantation: everolimus versus mycophenolate based immunosuppression regimen. Kidney Blood Press Res 2018;43(3):793–806. doi:10.1159/000489915

16. von Elm E, Altman DG, Egger M, et al; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61(4):344–349. doi:10.1016/j.jclinepi.2007.11.008

17. Ferri C, Giuggioli D, Guiducci S, et al; Italian Society for Rheumatology (SIR). Systemic sclerosis Progression INvestiGation (SPRING) Italian registry: demographic and clinico-serological features of the scleroderma spectrum. Clin Exp Rheumatol 2020;38 Suppl 125(3):40–47.

18. 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):2737–47.

19. Vassar M, Holzmann M. The retrospective chart review: important methodological considerations. J Educ Eval Health Prof 2013;10:12. doi:10.3352/jeehp.2013.10.12

20. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373–383. doi:10.1016/0021-9681(87)90171-8

21. Witman A, Beadles C, Liu Y, et al. Comparison group selection in the presence of rolling entry for health services research: rolling entry matching. Health Serv Res 2019;54(2):492–501. doi:10.1111/1475-6773.13086

22. Stuart EA, Lee BK, Leacy FP. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. J Clin Epidemiol 2013;66(Suppl 8):S84–S90.e1. doi:10.1016/j.jclinepi.2013.01.013

23. Kowal-Bielecka O, Landewé R, Avouac J, et al; EUSTAR Co-Authors. EULAR recommendations for the treatment of systemic sclerosis: a report from the EULAR Scleroderma Trials and Research group (EUSTAR). Ann Rheum Dis 2009;68(5):620–628. doi:10.1136/ard.2008.096677

24. Kowal-Bielecka O, Fransen J, Avouac J, et al; EUSTAR Coauthors. Update of EULAR recommendations for the treatment of systemic sclerosis. Ann Rheum Dis 2017;76(8):1327–1339. doi:10.1136/annrheumdis-2016-209909

25. Jaeger VK, Wirz EG, Allanore Y, et al; EUSTAR co-authors. Incidences and risk factors of organ manifestations in the early course of systemic sclerosis: a longitudinal EUSTAR study. PLoS One 2016;11(10):e0163894. doi:10.1371/journal.pone.0163894

26. Schoenfeld DA. Sample-size formula for the proportional-hazards regression model. Biometrics 1983;39(2):499–503. doi:10.2307/2531021

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1140–1148
DOI 10.1002/acr.70041
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

American College of Rheumatology
Empowering Rheumatology Professionals

Differential Item Functioning on the Patient Health Questionnaire 8 by Disease Subtype, Language, Sex, and Age Among People With Systemic Sclerosis: A Scleroderma Patient-Centered Intervention Network Cohort Study

Sophie Hu, BSc,[1,2] Marie-Eve Carrier, MSc,[1] Marie-Claude Geoffroy, PhD,[2,3] Meira Golberg, MMath,[1] Linda Kwakkenbos, PhD,[4,5] Susan J. Bartlett, PhD,[2,6] Catherine Fortuné,[7] Amy Gietzen,[8] Karen Gottesman, BA,[9] Geneviève Guillot, PDt,[10] Laura K. Hummers, MD,[11] Amanda Lawrie-Jones,[12,13] Vanessa L. Malcarne, PhD,[14] Michelle Richard, DSW,[15] Maureen Sauvé, BA,[16,17] Luc Mouthon, MD,[18] Andrea Benedetti, PhD,[2,6] and Brett D. Thombs, PhD,[1,2] on behalf of the SPIN Investigators

Objective. Somatic items used in depression assessments can potentially overlap with symptoms related to physical illness, including systemic sclerosis (SSc). No studies have looked at whether somatic depression items may be influenced by diffuse versus limited SSc disease subtypes, which are associated with varying degrees of symptom presentation. The objective of this study was to evaluate differential item functioning (DIF) in items of the 8-item Patient Health Questionnaire (PHQ-8) across SSc subtypes. We also assessed the PHQ-8 for DIF across language (English and French), sex, and age.

Methods. Participants enrolled in the Scleroderma Patient-Centered Intervention Network Cohort who completed the PHQ-8 at enrollment between April 2014 and October 2020 were included. Confirmatory factor analysis (CFA) was used to evaluate the unidimensional structure of the PHQ-8, and DIF analyses based on SSc subtype, language, sex, and age were conducted using Multiple Indicators Multiple Causes models.

Results. In total, 2,191 participants were included. CFA with several covarying error terms supported a one-factor structure for the PHQ-8 (Tucker–Lewis Index = 0.99, Comparative Fit Index = 0.98, Root Mean Square Error of Approximation = 0.08). We did not identify statistically significant DIF based on SSc subtype. Statistically significant DIF was found in one item for language, one item for sex, and two items for age. However, the effect of DIF on overall PHQ-8 scores was negligeable in all cases.

Conclusion. We did not find evidence that the PHQ-8 performs differently across SSc subtypes, language of administration, sex, and age groups.

INTRODUCTION

The Patient Health Questionnaire 9 (PHQ-9) is commonly used to assess depressive symptom severity in research and clinical care and, in some medical settings, to screen for depression.[1–4] The PHQ-9 contains nine items that reflect both psychological and somatic symptoms of depression, which align with the nine-symptom criteria for a major depressive episode

No funder had any role in the study design, collection, analysis and interpretation of the data, writing of the report, or in the decision to submit the paper for publication.

Ms Hu's work was supported by a Canadian Institutes of Health (CIHR) Canadian Graduate Scholarship Master's award, and Drs Geoffroy and Thombs' work was supported by Canada Research Chairs. The Scleroderma Patient-Centered Intervention Network Cohort has received funding from the CIHR; the Arthritis Society, the Lady Davis Institute for Medical Research of the Jewish General Hospital, Montreal, Quebec, Canada; the Jewish General Hospital Foundation, Montreal, Quebec, Canada; McGill University, Montreal, Quebec, Canada; the Scleroderma Society of Ontario; Scleroderma Canada; Sclérodermie Québec; Scleroderma Manitoba; Scleroderma Atlantic; the Scleroderma Association of British Columbia; Scleroderma Association of Saskatchewan; Scleroderma Australia; Scleroderma New South Wales; Scleroderma Victoria; and the Scleroderma Foundation of California.

[1]Lady Davis Institute of the Jewish General Hospital, Montreal, Quebec, Canada; [2]McGill University, Montreal, Quebec, Canada; [3]Douglas Mental Health University Institute, Montreal, Quebec, Canada; [4]Radboud University Behavioural Science Institute, Nijmegen, the Netherlands; [5]Radboud University Medical Center, Nijmegen, the Netherlands; [6]Research Institute, McGill University Health Centre, Montreal, Quebec, Canada; [7]Ottawa Scleroderma Support Group, Ottawa, Ontario, Canada; [8]Steffens Scleroderma Foundation, Albany, New York; [9]National Scleroderma Foundation, Los Angeles, California; [10]Sclérodermie Québec, Longueuil, Quebec, Canada; [11]Johns Hopkins University School of Medicine, Baltimore, Maryland; [12]Scleroderma Australia, Melbourne, Victoria, Australia; [13]Scleroderma Victoria, Melbourne, Victoria, Australia; [14]San Diego State University, San Diego, California; [15]Scleroderma Atlantic, Halifax, Nova Scotia, Canada; [16]Scleroderma Society of Ontario, Hamilton, Ontario, Canada; [17]Scleroderma Canada, Hamilton, Ontario, Canada; [18]Hôpital Cochin, Paris, France.

SIGNIFICANCE & INNOVATIONS

- Different disease presentation between people with diffuse systemic sclerosis (SSc) and those with limited SSc could influence the reporting of depression symptoms that overlap with disease symptoms (eg, fatigue, gastrointestinal symptoms, sleep problems).
- This was the first study to evaluate differential item functioning (DIF) in items of the 8-item Patient Health Questionnaire (PHQ-8), a commonly used depression symptom measure, based on disease subtype (diffuse or limited), language of administration (English or French), sex, and age in SSc.
- No statistically significant DIF was detected for PHQ-8 items between people with diffuse SSc and those with limited SSc, and the magnitude of DIF was negligeable for language, sex, and age.
- The PHQ-8 can be used to assess depression symptoms in people with SSc regardless of disease subtype, language, sex, and age.

according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.[5] The eight-item version (PHQ-8), which is identical to the PHQ-9 except for excluding an item on thoughts of suicide or self-harm, performs equivalently to the PHQ-9.[6] There is concern, however, that items on the PHQ-8 and PHQ-9 that assess somatic symptoms of depression (eg, fatigue) overlap with symptoms of many physical health conditions.[7–9] This could artifactually inflate scores and result in overestimating depression symptom severity or overidentifying people as possibly having depression.

Differential item functioning (DIF) analyses assess the degree that items in measurement tools reflect the construct being measured without being influenced by external factors that are not part of the construct.[10] DIF related to somatic symptoms on the PHQ-8 or PHQ-9 would be present if people with medical conditions or people with greater medical disease severity score higher on somatic symptoms than other people with similar levels of depression due to the overlap of somatic symptoms between mental and physical health conditions. We identified three studies that have assessed DIF in PHQ-8 or PHQ-9 items by comparing individuals with a medical condition, including rheumatoid arthritis (total N in study = 3,512),[11] multiple sclerosis (total N in study = 555),[12] and spinal cord injury (total N in study = 6,694),[13] to healthy comparators; none found item-level DIF that was statistically significant and of a magnitude that was clinically meaningful. We identified two other studies that have compared PHQ-9 responses among individuals with the same medical condition but different levels of medical disease severity or symptoms, one conducted with patients with cancer (N = 251)[14] and one with people with multiple sclerosis (N = 513).[15] Neither found significant DIF related to severity or symptom levels.

One additional study did not examine item-level DIF but compared PHQ-9 somatic symptom factor scores between people with systemic sclerosis (SSc; also known as scleroderma) and a sample of people not known to have a chronic medical condition (total N in study = 1,524). The study found that total somatic symptom scores (four items) were higher (standardized mean difference [SMD] = 0.55) among people with SSc, controlling for cognitive and affective scores (five items), sex, and age.[16]

SSc is a rare, chronic autoimmune disease involving microvascular damage and characterized by fibrosis of the skin and other organs, including the lungs, gastrointestinal tract, kidneys, and heart.[17–19] Disease onset typically peaks around age 50 years, and over 80% of people with the disease are women.[19] SSc can be categorized into two main cutaneous subtypes, diffuse and limited. Diffuse SSc is associated with skin thickening in larger areas of the body, including the trunk and the upper arms and legs, whereas in limited SSc, skin involvement is restricted to distal limbs (eg, hands and forearms) and the face.[18,19] Approximately 40% of people with SSc are classified with diffuse SSc.[20] Common SSc symptoms include skin thickening, Raynaud phenomenon, difficulty breathing, limitations in hand mobility, and three symptoms measured by PHQ-9 and PHQ-8 somatic items: fatigue, gastrointestinal issues, which can present as changes in appetite, and sleep problems.[17–19]

Although a previous study found that PHQ-9 and PHQ-8 somatic items may potentially be biased for people with SSc compared to healthy individuals,[16] it is not clear whether scores within a sample of people with SSc might be influenced by disease subtype, which is associated with differences in symptom presentation. Furthermore, little is known about whether other factors might influence PHQ-9 or PHQ-8 responses in SSc. This includes language of administration, which could affect the interpretation of questionnaire items, as well as sex or gender and age, which could also be associated with differences in symptom presentation. The primary objective of this study was to evaluate potential DIF in PHQ-8 items related to disease subtype (diffuse vs limited) in people with SSc and analyze the magnitude of any detected DIF. The secondary objective was to evaluate potential DIF related to language of administration (English vs French), sex (male vs female), and age.

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.70041).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.70041.

Address correspondence via email to Brett D. Thombs, PhD, at brett.thombs@mcgill.ca.

Submitted for publication June 2, 2025; accepted in revised form December 12, 2025.

PATIENTS AND METHODS

We conducted a cross-sectional analysis of baseline data collected at the time of enrollment in the Scleroderma Patient-Centered Intervention Network (SPIN) Cohort.[21] We first evaluated the PHQ-8 factor structure and then the DIF by SSc disease subtype, language, sex, and age. A protocol was published online before study initiation (https://osf.io/59pdb). We reported the study consistent with the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) reporting guideline for studies on properties of patient-reported outcome measures.[22] Studies that use SPIN Cohort data have overlapping methods. Thus, we reported results consistent with guidance from the Text Recycling Research Project.[23]

Participants and procedure. The SPIN Cohort collects longitudinal data on patient-reported outcomes from people living with SSc from 54 sites in Canada, the United States, the United Kingdom, France, Mexico, Spain, and Australia. Eligibility to participate in the SPIN Cohort requires being classified as having SSc based on 2013 American College of Rheumatology/EULAR criteria[24] confirmed by a SPIN physician; being ≥18 years; having fluency in English, French, or Spanish; and having access to a computer or tablet with internet access. The SPIN Cohort is a convenience sample. SPIN Cohort participants are recruited during regular medical visits at SPIN recruitment sites,[21] and written informed consent is obtained. A medical form is submitted online by site personnel to enroll participants. Once online registration is completed, participants are sent an automated welcome email with instructions on how to activate their SPIN account and complete SPIN Cohort measures online. Cohort participants complete online assessments upon enrollment and at three-month intervals.

The PHQ-8 was administered as a routine SPIN Cohort measure from cohort inception in April 2014 until October 2020. The present study used baseline assessment data for participants with complete PHQ-8 item-level data in English or French. Participants with missing baseline PHQ-8 data or DIF analysis variables were excluded.

The SPIN Cohort study was approved by the Research Ethics Committee of the Centre intégré universitaire de santé et de services sociaux (CIUSSS) du Centre-Ouest-de-l'Île-de-Montréal (#MP-05-2013-150) and by the ethics committees of all recruiting sites. A full list of the SPIN Investigators is included in Appendix A.

Measures. Marital status, self-reported race or ethnicity from a set of categories, country of residence, education, and language variables were collected from participants at baseline upon enrollment into the SPIN Cohort. Baseline data also included physician-reported medical data including age, sex, SSc subtype, time since initial onset of non-Raynaud phenomenon symptoms, presence of overlap syndromes (rheumatoid arthritis, Sjögren disease, autoimmune thyroid disease, systemic lupus erythematosus, and idiopathic inflammatory myositis), and SSc disease manifestations (digital ulcers, skin involvement, current tendon friction rubs, presence of joint contractures, presence of gastrointestinal symptoms, presence of interstitial lung disease, presence of pulmonary arterial hypertension, presence of biliary cholangitis, and history of renal crisis).

PHQ-8. The PHQ-8 consists of eight items that measure depression symptoms over the last two weeks. Items are scored on a 4-point scale ranging from 0 (not at all) to 3 (nearly every day) with higher scores (range 0–24) indicating more depression symptoms. The PHQ-8 has been shown to be equivalent to the PHQ-9,[6] and both the PHQ-9 and PHQ-8 have been shown to be reliable and valid measures of depression symptoms in SSc.[25,26] The PHQ-8 is available in English and French,[27] as well as in Spanish.[28] A systematic review found that 19 studies (58% of included studies) supported a one-factor structure for the PHQ-9,[29] and a previous study that performed a DIF analysis on a sample of people with SSc used a one-factor model.[27] Aside from being a common depression screening tool, the PHQ-9 is commonly used as a continuous patient-reported outcome measure for assessing depression symptoms in research.[1,30,31]

Statistical analysis. Descriptive statistics were calculated for the entire sample and, separately, for participants with diffuse SSc and limited SSc as mean (SD) for continuous variables and as frequencies and percentages for categorical variables.

Confirmatory factor analysis. First, we performed a confirmatory factor analysis (CFA) to evaluate whether a single-factor structure would be appropriate for DIF analyses. Item responses for the PHQ-8 were modeled as ordinal Likert data using a weighted least squares estimator, a diagonal weight matrix, and robust SEs. We assessed model fit using the Tucker–Lewis Index (TLI), the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA). A good model fit is indicated by a TLI and a CFI of ≥0.95 and an RMSEA of ≤0.06.[32] However, a TLI of ≥0.90 and an RMSEA of ≤0.08 are also often used to indicate an acceptable model fit.[33] We used modification indices to identify correlated measurement errors between pairs of items for which model fit would improve if we allowed error estimates to covary, which sometimes occurs due to similar item content or wording for instance.[34] In cases in which there was a clear theoretical rationale for why item errors covaried (eg, similar wording, content), we incorporated a correlated covariance term for the item pair into the model.

DIF analysis. We performed DIF analyses using the Multiple Indicators Multiple Causes (MIMIC) model. DIF analyses compare patterns of item responses in subgroups and test whether individuals with similar levels of a latent construct respond to each item similarly, regardless of subgroup status. A simulation study that varied sample sizes of focal and reference groups, number of

items in a scale, amount of DIF, and floor effects showed that DIF analyses are accurate for very short scales (eg, five items) with a sufficiently large sample size.[35] MIMIC models for DIF assessment are based on structural equation models, in which subgroup variables (SSc subtype = diffuse vs limited or sine, language = English vs French, sex = male vs female, age = ≤60 years vs >60 years) are added to the basic CFA model as observed variables. Thus, the base MIMIC model consists of the CFA factor model with the additional regression of the latent factor on subgroups to control for subgroup differences on the level of the latent factor. We then assessed potential DIF, one subgroup variable at a time. For each subgroup variable, we regressed PHQ-8 items separately, one at a time, on the subgroup variable. We identified potential DIF if there was a statistically significant link between item scores and the subgroup variable, after controlling by subgroups for latent depression factor levels. If DIF was detected in more than one item, the item with the largest magnitude of statistically significant DIF was considered to have DIF, and the link between the subtype group variable and that item was included in our model. We then repeated this procedure with the remaining items, separately regressing each of them one at a time on the subtype group variable and including the item with the largest statistically significant DIF in our model. This procedure was repeated until none of the remaining items showed statistically significant DIF. We analyzed DIF for subtype, followed, in order, by language, sex, and age. In each analysis, we controlled for any DIF identified in previous analyses.

Once we identified all items with significant DIF, we assessed the potential magnitude of any DIF items. We did this by comparing the difference in latent depression factor levels between subgroup variables in the baseline unadjusted model and in the final model after adjusting for DIF.

All statistical analyses were conducted in the statistical software R.[36] CFA and MIMIC model fit were assessed using the lavaan package.[37]

Sample size calculation. Recommendations for CFA sample size vary. In the present study, we performed a one-factor CFA and multiple MIMIC models with eight indicators using a sample size of 2,191 participants. This number significantly exceeds the minimum sample size recommended by established recommendations and standards.[38,39]

RESULTS

Sample characteristics. Sociodemographic variables and disease characteristics are summarized in Table 1. In total, 2,191 SPIN Cohort participants had complete data for the PHQ-8, including 871 participants with diffuse SSc (39.8%), 1,916 female participants (87.4%), 1,810 White participants (82.7%), and 1,488 participants who completed the PHQ-8 in English (67.9%). Four SPIN Cohort participants had missing data for some PHQ-8 items and were excluded. Mean age was 55.0 (SD 12.7) years. Mean time since onset of first non-Raynaud phenomenon symptoms was 11.1 (SD 8.8) years. Mean PHQ-8 score was 6.4 (SD 5.5), including 6.8 (SD 5.6) for participants with diffuse SSc versus 6.1 (SD 5.3) for those with limited SSc or sine scleroderma. Histograms with item response distributions are shown in Supplementary Material Figures 1–8.

CFA. The initial CFA model fit for a one-factor model was suboptimal (TLI = 0.96, CFI = 0.97, RMSEA = 0.12). Based on modification indices, we freed error covariance to correlate for items with similar content, including items 3 (sleep disturbances) and 4 (lack of energy or fatigue), items 2 (depressed mood) and 6 (negative feelings about self), and items 7 (difficulty concentrating) and 8 (psychomotor disturbances). The correlation between these pairs of items may be explained by the similarity of their content. For example, items 3 and 4 are both associated with tiredness, items 2 and 6 both describe a negative affective state, and items 7 and 8 can both be linked to dysregulation in the body. This resulted in a good model fit (TLI = 0.99, CFI = 0.98, RMSEA = 0.08). See Table 2.

DIF analysis. The MIMIC model before DIF adjustment, which included the regression of the latent depression factor on SSc subtype, language, sex, and age, demonstrated good fit (TLI = 0.99, CFI = 0.99, RMSEA = 0.05; Table 2). The pre-DIF adjustment and DIF-adjusted MIMIC models for SSc subtype, language, sex, and age are presented in Table 2.

We did not identify any items with statistically significant DIF related to SSc subtype. We identified one item with DIF related to language, one item with DIF related to sex, and two items with DIF related to age. French-language participants scored higher than expected on item 2 (depressed mood; β = 0.21, 95% confidence interval [CI] 0.13–0.28) compared to English-language participants, male participants scored lower than expected on item 3 (sleep disturbances; β = –0.13, 95% CI –0.24 to –0.01) compared to female participants, and participants older than 60 years scored lower than expected on item 6 (negative feelings about self; β = –0.17, 95% CI –0.25 to –0.08) and on item 7 (difficulty concentrating; β = –0.09, 95% CI –0.18 to –0.01) compared to participants aged 60 years and younger.

Differences in latent depression factor levels were similar in pre-DIF adjustment and DIF-adjusted models for language (French/English; unadjusted SMD = 0.10, 95% CI 0.00–0.20; DIF-adjusted SMD = 0.07, 95% CI –0.04 to 0.16), sex (male/female; unadjusted SMD = –0.10, 95% CI –0.24 to 0.05; DIF-adjusted SMD = –0.08, 95% CI –0.22 to 0.07), and age (>60 years/≤60 years; unadjusted SMD = –0.38, 95% CI –0.48 to –0.28; DIF-adjusted SMD = –0.34, 95% CI –0.44 to –0.24).

Table 1. Sociodemographic and disease characteristics in 2,191 patients*

	Full sample (N = 2,191)		Diffuse SSc (n = 871)		Limited SSc (n = 1,320)	
	n[a]	Mean ± SD or n (%)	n	Mean ± SD or n (%)	n	Mean ± SD or n (%)
Sociodemographic variables						
Age, y	2,191	55.0 ± 12.7	871	51.9 ± 12.5	1,320	57.0 ± 12.4
18–60		1,438 (65.6)		656 (75.3)		782 (59.2)
61+		753 (34.4)		215 (24.7)		538 (40.8)
Sex (female)	2,191	1,916 (87.4)	871	741 (85.1)	1,320	1,175 (89.0)
Language (English)	2,191	1,488 (67.9)	871	636 (73.0)	1,320	852 (64.5)
Race or ethnicity (White)	2,189	1,810 (82.7)	870	653 (75.1)	1,319	1,157 (87.7)
Education, y	2,187	14.9 ± 3.7	870	15.1 ± 3.5	1,317	14.8 ± 3.8
Marital status	2,191		871		1,320	
Married or living as married		1,552 (70.8)		597 (68.5)		955 (72.3)
Single		284 (13.0)		145 (16.6)		139 (10.5)
Divorced or separated		257 (11.7)		105 (12.1)		152 (11.5)
Widowed		98 (4.5)		24 (2.8)		74 (5.6)
Country	2,191		871		1,320	
United States		792 (36.1)		378 (43.4)		414 (31.4)
France		587 (26.8)		203 (23.3)		384 (29.1)
Canada		534 (24.4)		187 (21.5)		347 (26.3)
United Kingdom		237 (10.8)		95 (10.9)		142 (10.8)
Australia, Spain, Mexico		41 (1.9)		8 (0.9)		33 (2.5)
Disease characteristics						
Disease subtype	2,191					
Limited or sine[b]		1,320 (60.2)		0 (0)		1,320 (100)
Diffuse		871 (39.8)		871 (100)		0 (0)
Time since onset of first non-Raynaud symptom, y	2,010	11.1 ± 8.8	809	8.9 ± 7.5	1,201	12.6 ± 9.3
Gastrointestinal involvement	2,170	1,904 (87.7)	857	760 (88.7)	1,313	1,144 (87.1)
Digital ulcers	2,100	359 (17.1)	827	222 (26.8)	1,273	137 (10.8)
Tendon friction rubs	1,923		746		1,177	
Current		229 (11.9)		110 (14.7)		119 (10.1)
Past		221 (11.5)		181 (24.3)		40 (3.4)
Never		1,473 (76.6)		455 (61.0)		1,018 (86.5)
Small joint contractures	2,068		817		1,251	
None or mild		1,531 (74.0)		465 (56.9)		1,066 (85.2)
Moderate		389 (18.8)		248 (30.4)		141 (11.3)
Severe		148 (7.2)		104 (12.7)		44 (3.5)
Large joint contractures	2,027		804		1,223	
None or mild		1,768 (87.2)		633 (78.7)		1,135 (92.8)
Moderate		187 (9.2)		131 (16.3)		56 (4.6)
Severe		72 (3.6)		40 (5.0)		32 (2.6)
History of renal crisis	2,158	86 (4.0)	852	66 (7.7)	1,306	20 (1.5)
Interstitial lung disease	2,140	768 (35.9)	847	413 (48.8)	1,293	355 (27.5)
Pulmonary arterial hypertension	2,068	186 (9.0)	812	63 (7.8)	1,256	123 (9.8)
Primary biliary cholangitis	2,114	37 (1.8)	836	7 (0.8)	1,278	30 (2.3)
Overlap syndromes						
Systemic lupus erythematosus	2,137	65 (3.0)	843	22 (2.6)	1,294	43 (3.3)
Rheumatoid arthritis	2,133	123 (5.8)	842	59 (7.0)	1,291	64 (5.0)
Sjögren disease	2,099	165 (7.9)	830	50 (6.0)	1,269	115 (9.1)
Autoimmune thyroid disease	2,089	135 (6.5)	828	42 (5.1)	1,261	93 (7.4)
Idiopathic inflammatory myositis	2,129	110 (5.2)	835	60 (7.2)	1,294	50 (3.9)

* SSc, systemic sclerosis.
[a] Due to missing data, total sample size N < 2,191 for some characteristics.
[b] n with sine scleroderma = 63 (2.9% of total N).

DISCUSSION

We examined whether PHQ-8 item responses measured depression, free of influences from factors other than depression symptoms. We did not find any statistically significant DIF related to SSc disease subtype. We did identify statistically significant DIF in one item for language of administration, one item for sex, and two items for age. The magnitude of DIF on the overall latent depression factor estimate, however, was not meaningful in any comparisons, suggesting that total PHQ-8 scores are comparable across the subgroups that we examined.

Table 2. Patient Health Questionnaire 8 factor loadings and DIF estimates in unadjusted and DIF-adjusted models with DIF evaluations*

	Unadjusted model estimate (95% CI)[a]	SSc subtype model estimate (95% CI)	Language model estimate (95% CI)	Sex model estimate (95% CI)	Age model estimate (95% CI)
Item					
1. Loss of interest	0.88 (0.86–0.90)	0.88 (0.86–0.90)	0.88 (0.86–0.90)	0.88 (0.86–0.90)	0.88 (0.86–0.90)
2. Depressed mood	0.86 (0.84–0.88)	0.86 (0.84–0.87)	0.86 (0.84–0.87)	0.86 (0.84–0.88)	0.86 (0.84–0.88)
3. Sleep disturbances	0.66 (0.63–0.69)	0.66 (0.63–0.69)	0.66 (0.63–0.69)	0.66 (0.63–0.69)	0.66 (0.63–0.69)
4. Fatigue	0.78 (0.75–0.80)	0.78 (0.75–0.80)	0.78 (0.75–0.80)	0.78 (0.75–0.80)	0.78 (0.75–0.80)
5. Changes in appetite	0.74 (0.72–0.76)	0.74 (0.72–0.77)	0.74 (0.72–0.77)	0.74 (0.72–0.77)	0.74 (0.72–0.77)
6. Negative feelings about self	0.77 (0.75–0.80)	0.77 (0.75–0.80)	0.77 (0.75–0.80)	0.77 (0.75–0.80)	0.77 (0.74–0.80)
7. Difficulty concentrating	0.78 (0.75–0.80)	0.78 (0.75–0.81)	0.78 (0.75–0.81)	0.78 (0.75–0.80)	0.78 (0.75–0.80)
8. Psychomotor disturbances	0.70 (0.66–0.73)	0.70 (0.66–0.73)	0.70 (0.66–0.73)	0.70 (0.66–0.73)	0.70 (0.66–0.73)
Direct effect attributable to group					
2. Depressed mood	–	–	0.21 (0.13–0.28)	–	–
3. Sleep disturbances	–	–	–	−0.13 (−0.24 to −0.01)	–
6. Negative feelings about self	–	–	–	–	−0.17 (−0.25 to −0.08)
7. Difficulty concentrating	–	–	–	–	−0.09 (−0.18 to −0.01)
SMD on latent depression factor in unadjusted model[b]	–	0.07 (−0.02 to 0.17)	0.10 (0.00–0.20)	−0.10 (−0.24 to 0.05)	−0.38 (−0.48 to −0.28)
SMD on latent depression factor in DIF-adjusted model[b]	–	–	0.07 (−0.04 to 0.16)	−0.08 (−0.22 to 0.07)	−0.34 (−0.44 to −0.24)

* CI, confidence interval; DIF, differential item functioning; SMD, standardized mean difference; SSc, systemic sclerosis.
[a] The unadjusted model consisted of the CFA model with regressions on SSc subtype, language, sex, and age. Factor loadings in the confirmatory factor analysis model without regressions were all within 0.01 of the unadjusted model factor loadings.
[b] SMDs on latent depression factor were calculated for SSc subtype as diffuse-limited, for language as French-English, for sex as male - female, and for age as > 60 years-≤ 60 years.

To our knowledge, the present study is the first to evaluate DIF of PHQ-8 or PHQ-9 items based on disease subtype, age, and sex in a SSc population using MIMIC models. A previous study compared PHQ-9 somatic symptom scores (four items) between a SSc sample and a general population healthy sample, matched as closely as possible on cognitive and affective scores (five items), sex, and age.[16] Compared to healthy participants, participants with SSc had somatic symptom scores on average one point higher (1.4 points higher after adjusting for remaining age differences after matching). The authors concluded that PHQ-9 cutoffs used for the general population may not be suitable for people with SSc. This differed from results from our study as we did not find any significant association between SSc disease burden and responses to PHQ-8 items.

It is possible that results differed between the two studies because we analyzed items individually rather than as a factor group. We did not, however, find any items that differed significantly, even in a sample of over 2,000 participants, which would have enabled us to detect even very small differences. It is possible that the reason we did not find any disease subtype–related DIF is that, although many disease characteristics differ greatly by subtype, the PHQ-8 somatic items that we assessed reflect very common symptoms for people with SSc.[40,41] Among participants in the present study, 88% experienced gastrointestinal issues, and this was similar for participants with diffuse (89%) and limited SSc (87%). In the previous study, on the other hand, the authors compared people with SSc to a healthy sample without long-term physical conditions.[16]

Another possibility is that the lack of DIF in PHQ-8 somatic items could be related to the context in which items are presented. The somatic items of the PHQ-8 are presented as part of a depression symptom assessment, along with cognitive and mood-related depression symptom items, which could create a context effect.[42] Participants may have been aware that they were answering a depression assessment questionnaire and might therefore have underreported somatic symptoms that they attributed to their SSc or to overlapping syndromes rather than to depression. This could be tested. For example, a randomized controlled trial could be used to compare the scores of PHQ-8 somatic items in SSc when they are administered in their regular context, as part of the PHQ-8, versus when they are administered without the cognitive and affective items, which should remove participants' awareness that these are depression symptom items.

The findings of the present study have important implications for research. Although differences in physical disease burden may affect PHQ-9 somatic items when comparing a SSc population with a healthy population,[16] we did not find that this was the case when comparing people with different SSc disease subtypes. We further demonstrated the comparability of PHQ-8 depression scores across the English and French languages, sex, and age in SSc. These findings support the use of the PHQ-8 in people with SSc without concern about noncomparability of scores across SSc subgroups.

There are several strengths to our study, including its large sample size and the use of data from a multicenter, multinational

cohort, which allowed us to assess outcome measures in people with SSc in English and in French, as well as compare measurement properties with respect to SSc disease subtype, language of administration, age, and sex. Our study also has limitations to consider. First, the SPIN Cohort is a convenience sample, and SPIN participants completed all outcome measures online. Thus, we were unable to control for factors that may have influenced study participation such as technical difficulties. Nonetheless, characteristics of participants in the SPIN Cohort are comparable to participants in other large SSc cohorts, suggesting that our findings are likely generalizable.[20] Second, the examination of language-based DIF was limited to English- and French-language participants. Spanish-language participants from the SPIN Cohort were not included in this study due to their small sample size. As such, the generalizability of our results to other language populations is unknown. Third, our study only evaluated uniform DIF, and not nonuniform DIF, with the assumption that the relationship between item responses and grouping variables is constant. However, because PHQ-8 scores are calculated as a sum score, which considers the weight of all items equally, the calculation of scores would not be affected by any nonuniform DIF.

In summary, we found that PHQ-8 somatic items did not display DIF across SSc disease subtypes, indicating that the questionnaire performed comparably for people with diffuse and limited SSc. Although several items displayed statistically significant DIF across language of administration, participant sex, and participant age, the overall impact of DIF on PHQ-8 scores was negligeable. Therefore, our results support the use of the PHQ-8 to assess depression symptoms in SSc regardless of disease subtype, language (English or French), sex, or age.

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 Thombs 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. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606–613. doi:10.1046/j.1525-1497.2001.016009606.x
2. Moriarty AS, Gilbody S, McMillan D, et al. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. Gen Hosp Psychiatry 2015;37(6):567–576. doi:10.1016/j.genhosppsych.2015.06.012
3. Levis B, Benedetti A, Thombs BD; DEPRESSion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019;365:l1476. doi:10.1136/bmj.l1476
4. Levis B, Bhandari PM, Neupane D, et al; Depression Screening Data (DEPRESSD) PHQ Group. Data-driven cutoff selection for the Patient Health Questionnaire-9 depression screening tool. JAMA Netw Open 2024;7(11):e2429630. doi:10.1001/jamanetworkopen.2024.29630
5. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. American Psychiatric Publishing; 2013.
6. Wu Y, Levis B, Riehm KE, et al. Equivalency of the diagnostic accuracy of the PHQ-8 and PHQ-9: a systematic review and individual participant data meta-analysis. Psychol Med 2020;50(8):1368–1380. doi:10.1017/S0033291719001314
7. Koenig HG, George LK, Peterson BL, et al. Depression in medically ill hospitalized older adults: prevalence, characteristics, and course of symptoms according to six diagnostic schemes. Am J Psychiatry 1997;154(10):1376–1383. doi:10.1176/ajp.154.10.1376
8. Sørensenf C, Friis-Hasché E, Haghfelt T, et al. Postmyocardial infarction mortality in relation to depression: a systematic critical review. Psychother Psychosom 2005;74(2):69–80. doi:10.1159/000083165
9. von Ammon Cavanaugh S. Depression in the medically ill. Critical issues in diagnostic assessment. Psychosomatics 1995;36(1):48–59. doi:10.1016/S0033-3182(95)71707-8
10. Walker CM. What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. J Psychoeduc Assess 2011;29(4):364–376. doi:10.1177/0734282911406666
11. Hu J, Ward MM. Screening for depression in arthritis populations: an assessment of differential item functioning in three self-reported questionnaires. Qual Life Res 2017;26(9):2507–2517. doi:10.1007/s11136-017-1601-x
12. Marrie RA, Lix LM, Bolton JM, et al; CIHR Team in Defining the Burden and Managing the Effects of Immune-mediated Inflammatory Disease. Assessment of differential item functioning of the PHQ-9, HADS-D and PROMIS-depression scales in persons with and without multiple sclerosis. J Psychosom Res 2023;172:111415. doi:10.1016/j.jpsychores.2023.111415
13. Cook KF, Kallen MA, Bombardier C, et al. Do measures of depressive symptoms function differently in people with spinal cord injury versus primary care patients: the CES-D, PHQ-9, and PROMIS®-D. Qual Life Res 2017;26(1):139–148. doi:10.1007/s11136-016-1363-x
14. Jones SMW, Ludman EJ, McCorkle R, et al. A differential item function analysis of somatic symptoms of depression in people with cancer. J Affect Disord 2015;170:131–137. doi:10.1016/j.jad.2014.09.002
15. Jones SMW, Salem R, Amtmann D. Somatic symptoms of depression and anxiety in people with multiple sclerosis. Int J MS Care 2018;20(3):145–152. doi:10.7224/1537-2073.2017-069
16. Leavens A, Patten SB, Hudson M, et al; Canadian Scleroderma Research Group. Influence of somatic symptoms on Patient Health Questionnaire-9 depression scores among patients with systemic sclerosis compared to a healthy general population sample. Arthritis Care Res (Hoboken) 2012;64(8):1195–1201. doi:10.1002/acr.21675
17. Allanore Y, Simms R, Distler O, et al. Systemic sclerosis. Nat Rev Dis Primers 2015;1(1):15002. doi:10.1038/nrdp.2015.2
18. Volkmann ER, Andréasson K, Smith V. Systemic sclerosis. Lancet 2023;401(10373):304–318. doi:10.1016/S0140-6736(22)01692-0
19. Denton CP, Khanna D. Systemic sclerosis. Lancet 2017;390(10103):1685–1699. doi:10.1016/S0140-6736(17)30933-9
20. Dougherty DH, Kwakkenbos L, Carrier ME, et al; SPIN Investigators. The Scleroderma Patient-Centered Intervention Network Cohort: baseline clinical features and comparison with other large scleroderma cohorts. Rheumatology (Oxford) 2018;57(9):1623–1631. doi:10.1093/rheumatology/key139

21. The Scleroderma Patient-centered Intervention Network. 2024. Accessed November 15, 2025. https://www.spinsclero.com/
22. Gagnier JJ, Lai J, Mokkink LB, et al. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res 2021;30(8):2197–2218. doi:10.1007/s11136-021-02822-4
23. Hall S, Moskovitz C, Pemberton M. Best Practices for Researchers. Text Recycling Research Project. 2021. Accessed November 15, 2025 Nov 15. https://textrecycling.org/resources/best-practices-for-researchers/
24. 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. doi:10.1002/art.38098
25. Milette K, Hudson M, Baron M, et al; Canadian Scleroderma Research Group. Comparison of the PHQ-9 and CES-D depression scales in systemic sclerosis: internal consistency reliability, convergent validity and clinical correlates. Rheumatology (Oxford) 2010;49(4):789–796. doi:10.1093/rheumatology/kep443
26. Mattsson M, Sandqvist G, Hesselstrand R, et al. Validity and reliability of the Patient Health Questionnaire-8 in Swedish for individuals with systemic sclerosis. Rheumatol Int 2020;40(10):1675–1687. doi:10.1007/s00296-020-04641-1
27. Arthurs E, Steele RJ, Hudson M, et al; (CSRG) Canadian Scleroderma Research Group. Are scores on English and French versions of the PHQ-9 comparable? An assessment of differential item functioning. PLoS One 2012;7(12):e52028. doi:10.1371/journal.pone.0052028
28. Gómez-Gómez I, Benítez I, Bellón J, et al. Utility of PHQ-2, PHQ-8 and PHQ-9 for detecting major depression in primary health care: a validation study in Spain. Psychol Med 2023;53(12):5625–5635. doi:10.1017/S0033291722002835
29. Lamela D, Soreira C, Matos P, et al. Systematic review of the factor structure and measurement invariance of the patient health questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. J Affect Disord 2020;276:220–233. doi:10.1016/j.jad.2020.06.066
30. Kroenke K, Spitzer RL, Williams JB, et al. The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010;32(4):345–359. doi:10.1016/j.genhosppsych.2010.03.006
31. International Alliance of Mental Health Research Funders. Driving the Adaption of Common Measures. 2025. Accessed November 15, 2025. https://iamhrf.org/common-measures/
32. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6(1):1–55. doi:10.1080/10705519909540118
33. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21(2):230–258. doi:10.1177/0049124192021002005
34. McDonald RP, Ho MHR. Principles and practice in reporting structural equation analyses. Psychol Methods 2002;7(1):64–82. doi:10.1037/1082-989X.7.1.64
35. Scott NW, Fayers PM, Aaronson NK, et al; EORTC Quality of Life Group; Quality of Life Cross-Cultural Meta-Analysis Group. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. J Clin Epidemiol 2009;62(3):288–295. doi:10.1016/j.jclinepi.2008.06.003
36. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2024. Accessed November 15, 2025. https://www.r-project.org/
37. Rosseel Y. lavaan: an R package for structural equation modeling. J Stat Softw 2012;48(2):1–36. doi:10.18637/jss.v048.i02
38. Wolf EJ, Harrington KM, Clark SL, et al. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013;73(6):913–934. doi:10.1177/0013164413495237
39. Mundfrom DJ, Shaw DG, Ke TL. Minimum sample size recommendations for conducting factor analyses. Int J Test 2005;5(2):159–168. doi:10.1207/s15327574ijt0502_4
40. Willems LM, Kwakkenbos L, Leite CC, et al. Frequency and impact of disease symptoms experienced by patients with systemic sclerosis from five European countries. Clin Exp Rheumatol 2014;32(6 suppl 86):S88–S93.
41. Bassel M, Hudson M, Taillefer SS, et al. Frequency and impact of symptoms experienced by patients with systemic sclerosis: results from a Canadian National Survey. Rheumatology (Oxford) 2011;50(4):762–767. doi:10.1093/rheumatology/keq310
42. Lee L, Krishnamurty P, Van Horn S. Question order effects. In: Cooper H, Coutanche MN, McMullen LM, et al, eds. APA Handbook of Research Methods in Psychology: Foundations, Planning, Measures, and Psychometrics. 2nd ed. American Psychological Association; 2023:277–296.

APPENDIX A: SPIN INVESTIGATORS

SPIN Investigators: Claire E. Adams (Lady Davis Institute for Medical Research, Montreal, Quebec, Canada), Marie Hudson (Department of Medicine, McGill University, Montreal, Quebec, Canada), Maureen D. Mayes (University of Texas McGovern School of Medicine, Houston, Texas), James Stempel (Scleroderma Chicago, Chicago, Illinois), Robyn K. Wojeck (Amgen Inc, Thousand Oaks, California), Christian Agard (Centre Hospitalier Universitaire - Hôtel-Dieu de Nantes, Nantes, France), Laurent Alric (CHU Rangueil, Toulouse, France), Marc André (Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France), Floryan Beaslay (CHU La Réunion, Saint-Denis, La Réunion, France), Elana J. Bernstein (Columbia University, New York, New York), Sabine Berthier (Centre Hospitalier Universitaire Dijon Bourgogne, Dijon, France); Lyne Bissonnette (Université de Sherbrooke, Sherbrooke, Quebec, Canada), Sophie Blaise (CHU Grenoble Alpes, Grenoble, France), Eva Bories (CHU Rangueil, Toulouse, France), Alessandra Bruns (Université de Sherbrooke, Sherbrooke, Quebec, Canada), Carlotta Cacciatore (Assistance Publique – Hôpitaux de Paris, Hôpital St-Louis, Paris, France), Patricia Carreira (Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain), Marion Casadevall (Assistance Publique - Hôpitaux de Paris, Hôpital Cochin, Paris, France), Benjamin Chaigne (Assistance Publique - Hôpitaux de Paris, Hôpital Cochin, Paris, France), Lorinda Chung (Stanford University, Stanford, California), Benjamin Crichi (Assistance Publique - Hôpitaux de Paris, Hôpital St-Louis, Paris, France), Thylbert Deltombe (CHU La Réunion, Saint-Denis, La Réunion, France), Christopher P. Denton (Royal Free London Hospital, London, United Kingdom), Tannvir Desroche (CHU La Réunion, Saint-Denis, La Réunion, France), Robyn Domsic (University of Pittsburgh, Pittsburgh, Pennsylvania), James V. Dunne (St. Paul's Hospital and University of British Columbia, Vancouver, British Columbia, Canada), Bertrand Dunogue (Assistance Publique - Hôpitaux de Paris, Hôpital Cochin, Paris, France), Regina Fare (Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain), Dominique Farge-Bancel (Assistance Publique - Hôpitaux de Paris, Hôpital St-Louis, Paris, France), Paul R. Fortin (CHU de Québec - Université Laval, Quebec, Quebec, Canada), Tracy Frech (Vanderbilt University, Nashville, Tennessee), Loraine Gauzère (CHU La Réunion, Saint-Denis, La Réunion, France), Anne Gerber (CHU La Réunion, Saint-Denis, La Réunion, France), Jessica K. Gordon (Hospital for Special Surgery, New York City, New York), Brigitte Granel-Rey (Université and Assistance Publique - Hôpitaux de Marseille, Hôpital Nord, Marseille, France), Aurélien Guffroy (Les Hôpitaux Universitaires de Strasbourg, Nouvel Hôpital Civil, Strasbourg, France), Geneviève Gyger (Jewish General Hospital and McGill University, Montreal, Quebec, Canada), Eric Hachulla (Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France),

Daphna Harel (New York University, New York, New York), Monique Hinchcliff (Yale School of Medicine, New Haven, Connecticut), Sabrina Hoa (Centre hospitalier de l'Université de Montréal – CHUM, Montreal, Quebec, Canada), Michael Hugues (Salford Royal NHS Foundation Trust, Salford, United Kingdom), Alena Ikic (CHU de Québec - Université Laval, Quebec, Quebec), Sindhu R. Johnson (Toronto Scleroderma Program, Mount Sinai Hospital, Toronto Western Hospital, and University of Toronto, Toronto, Ontario, Canada), Nader Khalidi (McMaster University, Hamilton, Ontario, Canada), Kimberly S. Lakin (Hospital for Special Surgery, New York City, New York), Marc Lambert (Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France), Maggie Larche (University of Calgary, Calgary, Alberta, Canada), David Launay (Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France), Yvonne C. Lee (Northwestern University, Chicago, Illinois), Paul Legendre (Centre Hospitalier du Mans, Le Mans, France), Catarina Leite (University of Minho, Braga, Portugal), Hélène Maillard (Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France), Nancy Maltez (University of Ottawa, Ottawa, Ontario, Canada), Joanne Manning (Salford Royal NHS Foundation Trust, Salford, United Kingdom), Isabelle Marie (CHU Rouen, Hôpital de Bois-Guillaume, Rouen, France), Maria Martin Lopez (Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain), Thierry Martin (Les Hôpitaux Universitaires de Strasbourg, Nouvel Hôpital Civil, Strasbourg, France), Ariel Masetto (Université de Sherbrooke, Sherbrooke, Quebec, Canada), Arsène Mekinian (Assistance Publique - Hôpitaux de Paris, Hôpital St-Antoine, Paris, France), Sheila Melchor Díaz (Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain), Morgane Mourguet (CHU Rangueil, Toulouse, France), Christelle Nguyen (Université Paris Descartes, Université de Paris, Paris, France, and Assistance Publique - Hôpitaux de Paris, Paris, France), Karen Nielsen (Scleroderma Society of Ontario, Hamilton, Ontario, Canada), Mandana Nikpour (St Vincent's Hospital and University of Melbourne, Melbourne, Victoria, Australia), Louis Olagne (Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France), Vincent Poindron (Les Hôpitaux Universitaires de Strasbourg, Nouvel Hôpital Civil, Strasbourg, France), Janet Pope (University of Western Ontario, London, Ontario, Canada), Susanna Proudman (Royal Adelaide Hospital and University of Adelaide, Adelaide, South Australia, Australia), Grégory Pugnet (CHU Rangueil, Toulouse, France), Loïc Raffray (CHU La Réunion, Saint-Denis, La Réunion, France), François Rannou (Université Paris Descartes, Université de Paris, Paris, France, and Assistance Publique - Hôpitaux de Paris, Paris, France), Alexis Régent (Assistance Publique - Hôpitaux de Paris, Hôpital Cochin, Paris, France), Frederic Renou (CHU La Réunion, Saint-Denis, La Réunion, France), Sébastien Rivière (Assistance Publique - Hôpitaux de Paris, Hôpital St-Antoine, Paris, France), David Robinson (University of Manitoba, Winnipeg, Manitoba, Canada), Esther Rodríguez Almazar (Servicio de Reumatologia del Hospital 12 de Octubre, Madrid, Spain), Tatiana Sofia Rodríguez-Reyna (Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico), Sophie Roux (Université de Sherbrooke, Sherbrooke, Quebec, Canada), Perrine Smets (Centre Hospitalier Universitaire Gabriel-Montpied, Clermont-Ferrand, France), Vincent Sobanski (Centre Hospitalier Régional Universitaire de Lille, Hôpital Claude Huriez, Lille, France), Robert F. Spiera (Hospital for Special Surgery, New York City, New York), Virginia Steen (Georgetown University, Washington, DC), Evelyn Sutton (Dalhousie University, Halifax, Nova Scotia, Canada), Carter Thorne (Southlake Regional Health Centre, Newmarket, Ontario, Canada), Damien Vagner (CHU La Réunion, Saint-Denis, La Réunion, France), John Varga (University of Michigan, Ann Arbor, Michigan), Pearce Wilcox (St. Paul's Hospital and University of British Columbia, Vancouver, British Columbia, Canada), Vanessa Cook (Jewish General Hospital, Montreal, Quebec, Canada), Cassidy Dal Santo (Jewish General Hospital, Montreal, Quebec, Canada), Monica D'Onofrio (Jewish General Hospital, Montreal, Quebec, Canada), and Elsa-Lynn Nassar (Jewish General Hospital, Montreal, Quebec, Canada).

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1149–1158
DOI 10.1002/acr.80010
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

American College of Rheumatology
Empowering Rheumatology Professionals

Use of a Molecular Signature Response Classifier to Inform Treatment Selection Improves Clinical Disease Activity Among Patients with Rheumatoid Arthritis Initiating a Biologic or Targeted Synthetic Disease-Modifying Antirheumatic Drug

Fenglong Xie,[1] Timothy Beukelman,[1] Nicholas P. McCormick,[1] and Jeffrey R. Curtis[1,2]

Objective. We assessed the effectiveness of PrismRA to improve clinical outcomes among patients with rheumatoid arthritis (RA) initiating treatment with a biologic or targeted synthetic disease-modifying antirheumatic drug (b/tsDMARD).

Methods. PrismRA incorporated 19 gene expression features and four clinical features to assess a patient's likelihood of inadequate response to tumor necrosis factor inhibitor (TNFi). PrismRA was assessed in a prospective, interventional cohort study of patients initiating treatment with a b/tsDMARD. PrismRA results were provided to treating rheumatologists and incorporated into the selection of TNFi vs non-TNFi for study treatment. External comparator patients were identified in a rheumatology provider electronic health records system and matched to PrismRA patients using propensity scores. All patients had moderate-high disease activity at baseline. The primary study outcome was achievement of minimal important difference (MID) in clinical disease activity index (CDAI) at 24 weeks. Last observation carried forward was used to impute missing data.

Results. There were 330 PrismRA cohort patients and 990 matched comparator patients. Key baseline patient characteristics were all well-balanced between cohorts. Study treatment selection was consistent with PrismRA results in 82% of PrismRA cohort patients. CDAI MID at 24 weeks was achieved by 63.0% of PrismRA patients and 42.4% of comparator patients (odds ratio 2.31, confidence interval 1.79–2.99).

Conclusion. PrismRA results informing selection of TNFi vs non-TNFi treatment was associated with better CDAI outcomes compared to matched external comparator patients. PrismRA testing helps fill the need for a precision medicine approach to more rapidly identify the most effective therapy for individual patients with RA.

INTRODUCTION

The current treatment paradigm for rheumatoid arthritis (RA) is a treat-to-target approach in which therapeutic agents are added or switched until patients achieve the desired level of improvement.[1,2] Treatment is typically initiated with conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), such as methotrexate.[1,2] If the treatment response to csDMARDs is inadequate, treatment is most commonly escalated with the initiation of tumor necrosis factor inhibitors (TNFi).[3] However, treatment with other biologic or targeted synthetic DMARDs (b/tsDMARDs) with different mechanisms of action may be considered equally appropriate.[1,2] Additionally, as many as 38% of patients will have an inadequate response to TNFi treatment,[4] and this inadequate response cannot be predicted for an individual patient with certainty. This leads to a trial-and-error approach in the treatment of each patient that may result in avoidable adverse events, increased health care costs, reduced patient confidence, and prolonged disease activity.[5–8] Thus, a precision medicine approach to identify the optimal selection of b/tsDMARDs for each individual patient is a major unmet need in the treatment of RA.

Supported by Scipher Medicine. Dr Curtis's work was supported by the NIH (grant P30-AR-072583).

[1]Foundation for Advancing Science, Technology, Education and Research, Birmingham, Alabama; [2]University of Alabama at Birmingham, Birmingham.

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.80010).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80010.

Address correspondence via email to Jeffrey R. Curtis, MD, MS, MPH, at jrcurtis@uabmc.edu.

Submitted for publication June 24, 2025; accepted in revised form January 8, 2026.

SIGNIFICANCE AND INNOVATIONS

- In the management of rheumatoid arthritis (RA), PrismRA is a molecular signature response classifier shown to predict inadequate response to tumor necrosis factor inhibitor (TNFi) treatment with a positive predictive value of 87.7%.
- This study compared clinical outcomes with and without the use of PrismRA to inform the selection of TNFi or non-TNFi among patients initiating new treatment.
- Use of PrismRA to guide treatment selection led to significantly better clinical disease activity index scores. PrismRA may help more rapidly identify the most effective therapy for individual patients with RA.

PrismRA is a molecular signature response classifier (MSRC) that incorporates 19 gene expression features and four clinical features (anti–cyclic citrullinated peptide antibody positivity, sex, body mass index, and patient global assessment [PtGA] of disease activity).[9] The PrismRA MSRC assesses the likelihood of inadequate response to TNFi therapy and has been shown to have a positive predictive value of 87.7% to identify inadequate response to TNFi, as defined by failure to achieve the American College of Rheumatology (ACR) criteria for 50% improvement[10] at six months of treatment.[11] In patients with a high predicted likelihood of inadequate response to TNFi, providers may choose to initiate treatment with a non-TNFi b/tsDMARD to increase the likelihood of achieving treat-to-target goals quicker. Accordingly, clinical outcomes in company-sponsored cohort studies where all patients undergo MSRC testing were significantly improved when the selection of b/tsDMARD was informed by PrismRA results.[12]

The performance of the PrismRA MSRC in wide-scale clinical use in the treatment of RA is less certain. To further assess the potential benefit of using PrismRA testing, we compared clinical outcomes from a prospective, single arm interventional cohort study of patients initiating a b/tsDMARD whose selection was informed by PrismRA to patients observed in the real-world setting.

METHODS

Two patient cohorts were analyzed in this study: participants in a prospective, unblinded interventional study of the MSRC (PrismRA cohort) and patients identified within the Excellence Network in Rheumatology to Innovate Care and High-impact research (ENRICH) receiving standard of care treatments (ie, an external comparator cohort). ENRICH contains real-world data from electronic health records (EHRs) from more than 700 private practice rheumatology providers throughout the United States.[13] All data analyzed in this study were de-identified and determined to be exempt from Institutional Review Board (IRB) oversight by Advarra (IRB no. 00000971).

Patients in the PrismRA cohort were prospectively enrolled in a clinical utility study of PrismRA for RA, entitled DRIVE (ClinicalTrials.gov ID NCT05161234), at 33 clinical sites in the United States from March 2022 through February 2023. There was no overlap between patients enrolled in DRIVE and previously reported studies of the PrismRA MSRC.[9,12,14–16] Key inclusion criteria were age ≥18 years old; diagnosis of RA per the 2010 ACR/EULAR classification criteria[17]; active, moderate-high disease activity with baseline clinical disease activity index (CDAI) >10[18]; two or more of each swollen and tender joints on the 28-joint count assessment at baseline; and eligible for treatment with any b/tsDMARD. Concomitant treatment with csDMARDs, nonsteroidal anti-inflammatory drugs, stable oral glucocorticoids (prednisone equivalent of ≤10 mg/day), and intra-articular or parenteral glucocorticoids up to two weeks before baseline were allowed. Patients were excluded if they had prior exposure to any non-TNFi b/tsDMARD. Baseline measurements were performed within seven weeks before b/tsDMARD initiation, at which time informed consent was obtained and the PrismRA test was performed. Details of the PrismRA MSRC test have been previously described.[9,11] The PrismRA test result included a continuous score from 1 to 25 that is classified into one of three tiered categories: no signal of TNFi inadequate response (score <10.6), high signal of TNFi inadequate response (score ≥10.6 and <18.5), and very high signal of TNFi inadequate response (score ≥18.5). The PrismRA result was provided to the treating physician, and the results were used to inform the treatment selection for all patients included in the study. Patients who did not initiate treatment with b/tsDMARD within seven weeks of the baseline PrismRA testing were excluded from analysis.

External comparator arm. Given that all patients in the DRIVE cohort were tested with the MSRC, the clinical outcomes for comparable patients with RA who were not tested are unknown. For that reason, we created a synthetic control arm, a methodology commonly used in clinical trials or prospective cohort studies where no comparator population is available.[19] Consistent with US Food and Drug Administration guidance for the creation of a synthetic control arm to closely mirror the inclusion criteria and other design features of the cohort study as much as possible in the synthetic control arm,[20] patients observed in the ENRICH EHR data warehouse were considered for inclusion in the study as comparators if they satisfied the same inclusion and exclusion criteria as the PrismRA cohort patients, with minor adaptations to accommodate the real-world data (eg, physician diagnosis of RA rather than application of the 2010 ACR/EULAR classification criteria). Measurement of baseline characteristics occurred at the most recent visit up to three months before b/tsDMARD initiation. Patients were required to have available baseline measurement of CDAI and repeated CDAI

measurement at week 12 or 24 following b/tsDMARD initiation, at minimum. The specific treatment selection was based upon the treating physician's clinical judgment and was not informed by PrismRA test results.

Among the pool of patients in the ENRICH data meeting the inclusion and exclusion, patients were matched with patients in the PrismRA cohort using a propensity score (PS) model with a variable matching ratio developed with the primary goal of achieving balance in baseline characteristics between the cohorts as assessed by achieving a standardized mean difference (SMD) of <0.1.[21] The PS model building process aimed to retain as many PrismRA cohort participants as possible while also including matched comparators with the most recent b/tsDMARD initiation dates. Covariates included in the PS logistic regression model were intended to mirror the data collection in the DRIVE cohort and included age, sex, race, ethnicity, body mass index, smoking status, comorbid conditions (lung disease; myocardial infarction; other cardiovascular conditions; stroke; hypertension; hip, leg, or spine facture; depression; diabetes mellitus; gastric ulcer; fibromyalgia), Rheumatic Disease Comorbidity Index,[22] RA disease duration, RA seropositivity, baseline CDAI, concurrent use of methotrexate, concurrent use of nonmethotrexate csDMARD (sulfasalazine, leflunomide, hydroxychloroquine), and daily oral glucocorticoid dose. Prior TNFi exposure (none vs any) was matched exactly between PrismRA cohort patients and external comparators. Some patients had incomplete (ie, missing) baseline covariates: comorbid condition assessment in the PrismRA cohort and smoking status in the external comparator cohort. The study treatment (ie, TNFi vs non-TNFi b/tsDMARD) was intentionally omitted from the PS matching process because treatment selection is the mediator of the potential benefit of PrismRA testing. Inclusion of nonlinear terms (eg, transformations, restricted cubic splines) and variable interactions were considered on the basis of balance diagnostics and subject matter expertise. Performance of the PS model was evaluated by the ability to achieve covariate balance between the PrismRA and external comparator cohorts.[23,24] In matched analyses, F tests for equality of variances for continuous variables were used.[21]

Patients in the external comparator cohort were PS matched with participants in the PrismRA cohort using R package MatchIt, using a stepwise approach with patients with complete PS covariate availability matched first, and patients with incomplete PS covariate availability in either the DRIVE cohort or EHR data matched thereafter.[25] Individuals were matched on the logit-transformed PS using an optimal nearest-neighbor without replacement method within calipers.[26] Optimal matching was used to minimize the total absolute difference in the logit PS across all matches.[27] Caliper widths were determined based on the SD of the logit of PS (eg, 0.2 × SD).[28,29] Each PrismRA cohort patient was matched with up to five (1:5) external comparator patients using a variable matching ratio, and the targeted total number of matched external comparator patients was approximately twice the size of PrismRA cohort.[30] The two matched cohorts (with and without missing covariate data) were then combined for analysis of clinical outcomes.

Given the variable ratio matching, a weighted average approach was applied to the matched external comparator cohort patients as follows: if a PrismRA cohort patient was matched to x comparators, then these matched comparators were assigned weight = N_c / xN_t, where N_t is the total number of matched PrismRA patients and N_c is the total number of matched comparator patients.[30] PS matching was used to estimate the average treatment effect for the treated, which provides an estimate of the treatment effect among patients who ultimately received the treatment under study (in this study, the PrismRA test).[31]

The b/tsDMARD study treatment selection (TNFi vs non-TNFi) was obtained from the study site for the PrismRA cohort and from medication prescription records for the external comparator cohort. For patients in the PrismRA cohort, the PrismRA test result was recorded, and treatment selection consistent with or not consistent with the test result was determined. For all patients in both cohorts, clinical outcomes were assessed at 12 and 24 weeks following study treatment initiation using the closest clinical assessment within a four-week window (ie, 8 to 16 weeks and 20 to 28 weeks, respectively).

Outcome. The primary study outcome was the proportion of patients who achieved a minimal important difference (MID) in CDAI of ≥6 (if baseline CDAI >10 and ≤22) or ≥12 (if baseline CDAI >22) at 24 weeks after study treatment initiation.[32] Secondary binary study clinical outcomes were the proportion of patients who achieved CDAI MID at 12 weeks after study treatment initiation and achievement of the following at 12 and 24 weeks after study treatment initiation: CDAI ≤10 (low disease activity)[18]; CDAI ≤2.8 (remission)[18]; meaningful reduction in PtGA of ≥10 on scale of 0 to 100 compared to baseline[33]; minimal clinically important difference (MCID) in Routine Assessment of Patient Index Data 3 (RAPID3) of ≥3.8 on scale of 0 to 30 compared to baseline[34]; and MCID in patient pain visual analog scale of ≥1.1 on the scale of 0–10.[35,36] Change in CDAI scores from baseline at 12 and 24 weeks following study treatment initiation was included as a secondary continuous study outcome.

No patients in either cohort were excluded for b/tsDMARD discontinuation or switching during follow-up. Last observation carried forward (LOCF) was used to impute missing clinical outcomes in the primary analysis of both cohorts; thus, all patients were included. All patients had either a 12- or 24-week CDAI available, and no patient had imputation of their baseline CDAI as the primary outcome. The effects of LOCF imputation were assessed by comparing the outcome measure results with LOCF versus complete case analysis (ie, excluding patients with missing results).

Statistical analysis. SMD and 95% confidence interval (CI) were generated to assess covariate balance between the PrismRA cohort and the weighted matched external comparator cohort. The dichotomous study clinical outcomes were evaluated using weighted logistic regression to produce odds ratios (ORs) with 95% CI. The continuous study clinical outcomes were evaluated using weighted linear regression to produce beta coefficients and 95% CI. Outcome regression models included direct adjustment for covariates that were not well-matched (ie, SMD >0.1).

Because prior TNFi use can strongly influence the choice of subsequent b/tsDMARD irrespective of PrismRA testing, a subgroup analysis was performed for patients naïve to TNFi (ie, all patients with prior TNFi use were excluded). An additional subgroup analysis was performed that excluded patients in the PrismRA cohort who received treatment with TNFi despite having a PrismRA result indicating high or very high signal of TNFi inadequate response (ie, received treatment not consistent with PrismRA result). The external comparator patients matched to excluded PrismRA patients were also excluded from this subgroup analysis.

RESULTS

There were 330 participants eligible from the PrismRA prospective cohort, and all 330 patients were matched to at least one external comparator patient and included in the analyses. A total of 2,632 potential external comparator patients who met study cohort inclusion and exclusion criteria were identified in the ENRICH data. A comparison of the baseline characteristics of all patients in the external comparator pool and the PrismRA cohort is shown in Supplemental Table 1. Following PS matching with the PrismRA cohort patients, 990 patients were included in the external comparator cohort. The study treatment initiation date of patients in the matched external comparator cohort ranged from January 2018 through June 2024.

The characteristics of PrismRA patients and weighted matched external comparator patients are presented in Table 1. Following PS matching and weighting, the baseline characteristics were similar between the PrismRA and external comparator cohort (Supplemental Figure 1). There was a small residual difference (SMD = 0.11) in the baseline glucocorticoid dose and body mass index categories in the complete and incomplete baseline covariate matched patients, respectively. The Rheumatic Disease Comorbidity Index was slightly higher in the PrismRA cohort (SMD = 0.13). There were also small differences in the frequency of uncommon comorbid conditions (lung disease and other cardiovascular disease). When combining the complete and incomplete baseline covariate matched cohorts, the median baseline CDAI in the PrismRA cohort was 28.8 and was 24.0 in the external comparator.

The study treatments prescribed following PrismRA testing and in the synthetic comparator arm are shown in Table 2, stratified by PrismRA results. Among all patients in the PrismRA cohort, 55.5% received treatment with TNFi. This is similar to the 57.9% of external comparator cohort patients who received treatment with TNFi. The majority of patients in the PrismRA cohort (82%) received treatment that was consistent with the PrismRA result. Overall, 59.7% of patients had a PrismRA signal of inadequate response. Among patients with no signal of TNFi inadequate response (n = 132; 40.0%), 93.9% received treatment with TNFi. Among patients with a high signal (n = 138; 41.8%) or very high signal (n = 59; 17.9%) of TNFi inadequate response, 28.3% and 33.9% of patients received treatment with TNFi, respectively. One patient's PrismRA test was not analyzable due to delayed receipt of the sample, and this patient was included in all analyses.

The proportions of patients achieving the binary study clinical outcomes and the corresponding ORs are shown in Table 3. The primary study outcome of a MID in the CDAI (ie, improvement >12 units for those starting in high disease activity, and >6 in those starting in moderate disease activity) at week 24 was achieved by 63.0% of patients in the PrismRA cohort compared to 42.4% of patients in the external comparator cohort (OR 2.31, 95% CI 1.79–2.99; $P < 0.001$). All secondary binary clinical outcomes that assessed the CDAI were also significantly associated with PrismRA ($P < 0.001$), ranging from an OR of 2.52 (95% CI 1.93–3.28) for CDAI low disease activity at week 24 to an OR of 41.1 (95% CI 15.9–148) for CDAI remission at week 24, although CDAI remission at week 24 was only achieved by 13.0% of patients in the PrismRA cohort.

Improvements in RAPID3 were significantly more frequent in the PrismRA cohort at week 12 (OR 1.33, 95% CI 1.00–1.75) and week 24 (OR 1.47, 95% CI 1.13–1.91). Improvements in PtGA were not significantly different in the PrismRA cohort compared to external comparator cohort at 12 or 24 weeks. Improvements in patient pain scale in the PrismRA cohort were significantly better at week 12 (OR 1.63, 95% CI 1.26–2.12) but not at week 24. The improvement in CDAI in the PrismRA cohort at week 12 (median 10.5, interquartile range [IQR] 0.0–19.0) and week 24 (median 13.0, IQR 4.0–22.0) was greater than the external comparator cohort (median 2.0 [IQR 0.0–11.5] and median 6.0 [IQR –0.5 to 15.0], respectively; $P < 0.001$ for both time points).

The subgroup analyses restricted to TNFi-naïve patients is shown in Table 4. The estimate of the primary study outcome of achievement of CDAI MID at week 24 was essentially unchanged (OR 2.44, 95% CI 1.80–3.31; $P < 0.001$). Similarly, the estimates for all of the binary secondary clinical outcomes were essentially unchanged from the results of the full cohort (Table 4).

The subgroup analyses that excluded 59 PrismRA cohort patients who initiated TNFi despite having a PrismRA result indicating a high or very high signature of inadequate response to TNFi therapies produced results very similar to the primary analysis that included all PrismRA patients (Table 5) at 12 and

Table 1. Characteristics of PrismRA patients and weighted matched comparator patients, stratified by complete versus incomplete baseline covariates*

Characteristic	Complete baseline covariates				Incomplete baseline covariates			
	Comparator (n = 765)	PrismRA (n = 255)	SMD	95% CI	Comparator (n = 225)	PrismRA (n = 75)	SMD	95% CI
Age, median (IQR), y	56.0 (47.0–64.0)	57.0 (47.8–65.3)	−0.04	−0.18 to 0.10	56.0 (47.0–65.0)	56.5 (46.5–66.0)	0.00	−0.26 to 0.26
Gender, n (%)			0.02	−0.12 to 0.16			0.08	−0.18 to 0.34
Female	619 (80.9)	204 (80.0)			187 (83.1)	60 (80.0)		
Male	146 (19.1)	51 (20.0)			38 (16.9)	15 (20.0)		
Race, n (%)			0.08	−0.06 to 0.22			0.07	−0.19 to 0.33
Black or African American	76 (9.9)	25 (9.8)			14 (6.2)	6 (8.0)		
White	588 (76.9)	203 (79.6)			182 (80.9)	60 (80.0)		
Other	101 (13.3)	27 (10.6)			29 (12.9)	9 (12.0)		
Ethnicity, n (%)			0.03	−0.11 to 0.17			0.05	−0.22 to 0.31
Hispanic or Latino	274 (35.8)	95 (37.3)			82 (36.4)	29 (38.7)		
Non-Hispanic or Latino	491 (64.2)	160 (62.7)			143 (63.6)	46 (61.3)		
BMI category, n (%)			0.03	−0.12 to 0.17			0.11	−0.15 to 0.38
Underweight	2 (0.2)	1 (0.4)			3 (1.3)	1 (1.3)		
Healthy weight	154 (20.1)	51 (20.0)			55 (24.4)	15 (20.0)		
Overweight	247 (32.2)	83 (32.5)			66 (29.3)	22 (29.3)		
Obesity	363 (47.5)	120 (47.1)			101 (44.9)	37 (49.3)		
Smoke, n (%)			0.08	−0.06 to 0.22	—	—	—	—
Current	73 (9.5)	30 (11.8)			—	—	—	—
Never	545 (71.2)	180 (70.6)			—	—	—	—
Past	148 (19.3)	45 (17.6)			—	—	—	—
Lung disease, n (%)	32 (4.2)	19 (7.5)	0.14	0.00 to 0.29	—	—	—	—
Myocardial infarction, n (%)	4 (0.5)	2 (0.8)	0.03	−0.11 to 0.17	—	—	—	—
Other cardiovascular disease, n (%)	38 (5.0)	19 (7.5)	0.11	−0.04 to 0.25	—	—	—	—
Stroke, n (%)	8 (1.1)	2 (0.8)	0.04	−0.11 to 0.18	—	—	—	—
Hypertension, n (%)	269 (35.1)	93 (36.5)	0.03	−0.11 to 0.17	—	—	—	—
Fractures, n (%)	29 (3.8)	13 (5.1)	0.07	−0.07 to 0.21	—	—	—	—
Depression, n (%)	106 (13.9)	35 (13.7)	0.00	−0.14 to 0.15	—	—	—	—
Diabetes, n (%)	114 (15.0)	45 (17.6)	0.07	−0.07 to 0.21	—	—	—	—
Ulcer, n (%)	113 (14.8)	35 (13.7)	0.03	−0.11 to 0.17	—	—	—	—
Fibromyalgia, n (%)	49 (6.4)	14 (5.5)	0.03	−0.11 to 0.18	—	—	—	—
RDCI, n (%)			0.13	−0.01 to 0.27	—	—	—	—
0	346 (45.2)	104 (40.8)			—	—	—	—
1	206 (27.0)	65 (25.5)			—	—	—	—
2	132 (17.3)	52 (20.4)			—	—	—	—
3	81 (10.6)	34 (13.3)			—	—	—	—
RA disease duration, n (%)			0.03	−0.11 to 0.17			0.09	−0.17 to 0.35
≤2	319 (41.8)	110 (43.1)			69 (30.7)	26 (34.7)		
2–5	170 (22.3)	54 (21.2)			59 (26.2)	19 (25.3)		
>5	275 (36.0)	91 (35.7)			97 (43.1)	30 (40.0)		
Seropositivity, n (%)	468 (61.1)	148 (58.0)	0.06	−0.08 to 0.20	159 (70.7)	53 (70.7)	0.00	−0.26 to 0.26
CDAI at baseline, median (IQR)	26.5 (19.0, 40.5)	29.5 (21.0, 39.0)	−0.06	−0.20 to 0.08	25.8 (17.5, 37.4)	27.0 (20.5, 37.1)	−0.09	−0.35 to 0.17

(Continued)

Table 1. (*Cont'd*)

Characteristic	Complete baseline covariates				Incomplete baseline covariates			
	Comparator (n = 765)	PrismRA (n = 255)	SMD	95% CI	Comparator (n = 225)	PrismRA (n = 75)	SMD	95% CI
Prior TNFi use, n (%)	183 (23.9)	61 (23.9)	0.00	−0.14 to 0.14	93 (41.3)	31 (41.3)	0.00	−0.26 to 0.26
Concurrent methotrexate use, n (%)	464 (60.7)	162 (63.5)	−0.06	−0.20 to 0.08	112 (49.8)	36 (48.0)	0.04	−0.23 to 0.30
Concurrent conventional DMARD use other than methotrexate, n (%)	238 (31.2)	79 (31.0)	0.00	−0.14 to 0.15	55 (24.4)	19 (25.3)	−0.02	−0.28 to 0.24
Hydroxychloroquine	155 (20.3)	54 (21.2)			30 (13.3)	13 (17.6)		
Leflunomide	107 (14.0)	27 (10.6)			25 (11.1)	5 (6.8)		
Sulfasalazine	59 (7.7)	14 (5.5)			13 (5.8)	1 (1.4)		
Daily glucocorticoids dose at index date, n (%)			0.11	−0.03 to 0.25			0.06	−0.20 to 0.32
None	614 (80.3)	195 (76.5)			171 (76.0)	56 (74.7)		
0–7.5 mg	101 (13.2)	43 (16.9)			42 (18.7)	14 (18.7)		
≥7.5 mg	50 (6.5)	17 (6.7)			12 (5.3)	5 (6.7)		

* BMI, body mass index; CDAI, Clinical Disease Activity Index; CI, confidence interval; DMARD, disease-modifying antirheumatic drug; IQR, interquartile range; RA, rheumatoid arthritis; RDCI, Rheumatic Disease Comorbidity Index; SMD, standardized mean difference; TNFi, tumor necrosis factor inhibitor.

Table 2. Study treatment received by PrismRA and external comparator patients*

Patients	b/tsDMARD initiated, n (%)	
	TNFi	Non-TNFi
All PrismRA (N = 330)	183 (55.5)	147 (44.5)
No signal of TNFi inadequate response (n = 132)	124 (93.9)	8 (6.1)
High signal of TNFi inadequate response (n = 138)	39 (28.3)	99 (71.7)
Very high signal of TNFi inadequate response (n = 59)	20 (33.9)	39 (66.1)
Sample not analyzable (n = 1)	0 (0)	1 (100)
All comparators (n = 990)	573 (57.9)	417 (42.1)

* b/tsDMARD, biologic or targeted synthetic disease-modifying antirheumatic drug; TNFi, tumor necrosis factor inhibitor.

24 weeks. The primary study outcome of achievement of CDAI MID at week 24 was significantly associated with the PrismRA testing compared to the external comparator cohort (OR 2.49, 95% CI 1.87–3.32; $P < 0.001$). Similar to the primary analyses, in the subgroup analyses, all outcomes of CDAI improvement were significantly better in the PrismRA cohort (Table 5).

Regarding the impact of LOCF imputation, the CDAI outcome at 24 weeks was missing in 25.5% of PrismRA and 41.5% of external comparator patients (Supplemental Table 2). Compared to a complete case analysis, the median LOCF 24-week CDAI score worsened from 10.5 to 14.0 in the PrismRA cohort and improved slightly from 17.5 to 17.0 in the external comparator. Similarly, in LOCF compared to complete case analysis the 24-week median PtGA, RAPID3, and pain scores were all worse in the PrismRA cohort and slightly improved or unchanged in the external comparator.

DISCUSSION

The availability and application of PrismRA results in the selection of b/tsDMARD in a prospective interventional cohort of patients with moderate-high RA disease activity resulted in a greater proportion of patients achieving CDAI MID at 24 weeks versus an external observational matched comparator cohort of highly similar patients who initiated b/tsDMARD treatment without PrismRA testing. All secondary CDAI-based outcomes were significantly better among patients in the PrismRA versus external comparator cohort as well. Meaningful improvements in RAPID3 at 12 and 24 weeks and patient pain scale at 12 weeks were also significantly more common in PrismRA versus external comparator patients. Restricting the analyses to TNFi-naïve patients produced nearly identical results.

The results of this study support the use of PrismRA MSRC testing to improve clinical outcomes in patients needing treatment with b/tsDMARD. Despite numerous interventional and observational comparative effectiveness studies among the different b/tsDMARDs showing comparable efficacy in large groups of patients, no clearly superior therapeutic agent or mechanism of action has been identified for individual patients with RA.[1,2] As a result, a trial-and-error strategy is used in the absence of a personalized medicine approach. Based on the results of this study showing a risk difference of 20.6% (63.0% in PrismRA minus 42.4% in external comparator) for the achievement of CDAI MID, the number of patients needed to test with the PrismRA MSRC that was associated with one additional patient achieving a CDAI MID was approximately 5 (1 / 0.206), representing an important step toward personalized medicine. Unfortunately, one potential barrier to the successful implementation of the PrismRA MSRC in clinical practice is the limited access to various b/tsDMARD that is routinely imposed by payors. Indeed, in a previously published study assessing the clinical utility of PrismRA testing, insurance-related restrictions were responsible for 62.0% of deviations from the treatment recommended by PrismRA results.[16]

Despite the prospective study being conducted predominantly in the post–COVID-19 pandemic era in which care might differ compared to the pre–COVID-19 or peri–COVID-19 era, treatment selection in the PrismRA cohort was consistent with

Table 3. Comparison of binary clinical outcomes at 12 and 24 weeks following study treatment initiation in PrismRA cohort patients versus the weighted external comparator cohort using LOCF imputation*

Outcome	PrismRA (n = 330), n (%)	Comparator (n = 990), n (%)	PrismRA vs comparator, OR (95% CI)	*P* value
CDAI MID at wk 24	208 (63.0)	420 (42.4)	2.31 (1.79–2.99)	<0.001
CDAI MID at wk 12	178 (53.9)	303 (30.6)	2.66 (2.06–3.43)	<0.001
CDAI remission (≤2.8) at wk 12	21 (6.4)	3 (0.3)	24.0 (7.99–108)	<0.001
CDAI low disease activity (≤10.0) at wk 12	106 (32.1)	154 (15.6)	2.57 (1.92–3.42)	<0.001
CDAI remission (≤ 2.8) at wk 24	43 (13.0)	4 (0.4)	41.1 (15.9–148)	<0.001
CDAI low disease activity (≤10.0) at wk 24	138 (41.8)	220 (22.2)	2.52 (1.93–3.28)	<0.001
RAPID3 improvement ≥3.8 at wk 12	99/315 (31.4)	241/941 (25.7)	1.33 (1.00–1.75)	0.046
RAPID3 improvement ≥3.8 at wk 24	131/315 (41.6)	307/941 (32.6)	1.47 (1.13–1.91)	0.004
PtGA improvement ≥10 at wk 12	137 (41.5)	432 (43.7)	0.92 (0.71–1.18)	0.5
PtGA improvement ≥10 at wk 24	169 (51.2)	488 (49.3)	1.08 (0.84–1.39)	0.5
Patient pain scale improvement ≥1.1 at wk 12	132 (40.0)	279/960 (29.0)	1.63 (1.26–2.12)	<0.001
Patient pain scale improvement ≥1.1 at wk 24	157 (47.6)	421/960 (43.9)	1.16 (0.90–1.49)	0.2

* CDAI, Clinical Disease Activity Index; CI, confidence interval; LOCF, last observation carried forward; MID, minimal important difference; OR, odds ratio; PtGA, patient global assessment; RAPID3, Routine Assessment of Patient Index Data 3.

Table 4. Comparison of binary clinical outcomes at 12 and 24 weeks following study treatment initiation in PrismRA cohort patients versus the weighted external comparator cohort using LOCF imputation, restricted to TNFi-naïve patients*

Outcome	PrismRA (n = 238), n (%)	Comparator (n = 714), n (%)	PrismRA versus comparator, OR (95% CI)	*P* value
CDAI MID at wk 24	154 (64.7)	307 (42.9)	2.44 (1.80–3.31)	<0.001
CDAI MID at wk 12	142 (59.7)	235 (32.9)	3.02 (2.23–4.09)	<0.001
CDAI remission (≤2.8) at wk 12	17 (7.1)	2 (0.3)	24.9 (7.35–142)	<0.001
CDAI low disease activity (≤10.0) at wk 12	83 (34.9)	118 (16.6)	2.70 (1.94–3.76)	<0.001
CDAI remission (≤2.8) at wk 24	35 (14.7)	3 (0.4)	40.9 (14.5–171)	<0.001
CDAI low disease activity (≤10.0) at wk 24	100 (42.0)	163 (22.8)	2.45 (1.80–3.35)	<0.001
RAPID3 improvement ≥3.8 at wk 12	76/228 (33.3)	176/682 (25.8)	1.44 (1.04–1.99)	0.028
RAPID3 improvement ≥3.8 at wk 24	97/228 (42.5)	223/682 (32.8)	1.52 (1.12–2.06)	0.008
PtGA improvement ≥10 at wk 12	105 (44.1)	319 (44.7)	0.98 (0.73–1.31)	0.9
PtGA improvement ≥10 at wk 24	123 (51.7)	357 (50.0)	1.07 (0.80–1.44)	0.7
Patient pain scale improvement ≥1.1 at wk 12	101 (42.4)	207/695 (29.8)	1.73 (1.28–2.35)	<0.001
Patient pain scale improvement ≥1.1 at wk 24	118 (49.6)	303/695 (43.6)	1.27 (0.95–1.71)	0.11

* CDAI, Clinical Disease Activity Index; CI, confidence interval; LOCF, last observation carried forward; MID, minimal important difference; OR, odds ratio; PtGA, patient global assessment; RAPID3, Routine Assessment of Patient Index Data 3; TNFi, tumor necrosis factor inhibitor.

past PrismRA results for 82% of patients in this study, only slightly higher than previously reported (70% to 76%).[14,15] The proportion of patients with a PrismRA signal of TNFi inadequate response (59.7%) was identical to that previously reported (59%),[15] and the proportion of patients achieving CDAI MID at week 24 was comparable or only slightly better than observed in a previously published study of PrismRA testing (63.0% vs 49.5% for PrismRA and 42.4% vs 32.8% for external comparators[15]).

The clinical responses to treatment observed in our study displayed some differences from other previously published large observational studies. For example, in an observational study of patients with baseline moderate-high RA disease activity within the CorEvitas registry in the United States, initiation of treatment with TNFi in combination with methotrexate resulted in CDAI MID at six months in 56.9% of biologic-naïve patients (n = 297); for patients with prior biologic use (n = 161), the result was 41.6%.[37] Considering that 27.8% of patients had prior TNFi use in the present study, disease activity responses in the PrismRA cohort were higher than and the external comparator cohort results were lower than from this report. A second observational study of patients with RA in the United States (64.7% prior biologic use; mean baseline CDAI 31.5) who initiated treatment with the TNFi golimumab reported achievement of CDAI low disease activity at six months in 36.2% and 29.5% of patients with (n = 420) and without (n = 265) concomitant methotrexate, respectively.[38] Again, the treatment responses from the present study were better in the PrismRA cohort (41.8% CDAI low disease activity) and worse in the external comparator cohort (22.2% CDAI low disease activity) compared to the golimumab study. In contrast, a different analysis of biologic-naïve patients with RA in CorEvitas reported achievement of CDAI MID at 12 months in 45.5% of patients initiating TNFi (n = 2,372) and 45.0% of patients initiating non-TNFi b/tsDMARD (n = 593)[39]; these results are more similar to the response observed in the external comparator cohort of the present study.

This study used an external comparator of patients observed in a large database of rheumatology EHR records rather than a parallel arm randomized design. However, the eligibility criteria used to identify the large pool of potential comparator patients were highly similar to the PrismRA prospective cohort eligibility criteria, and subsequent PS-based matching was performed to successfully balance important baseline patient characteristics between the two cohorts, such as demographics, RA disease duration and seropositivity, prior and concurrent medication use, and baseline disease activity. It seems unlikely that any

Table 5. Comparison of binary clinical outcomes at 12 and 24 weeks following study treatment initiation in PrismRA cohort patients versus the weighted external comparator cohort using LOCF imputation, excluding patients who initiated TNFi despite having a PrismRA result indicating a molecular signature of inadequate response to TNFi and their matched comparators*

Outcome	PrismRA (n = 271), n (%)	Comparator (n = 813), n (%)	PrismRA versus comparator, OR (95% CI)	*P* value
CDAI MID at wk 24	176 (64.9)	347 (42.7)	2.49 (1.87–3.32)	<0.001
CDAI MID at wk 12	143 (52.8)	251 (30.9)	2.50 (1.89–3.31)	<0.001
CDAI remission (≤2.8) at wk 12	15 (5.5)	3 (0.3)	17.0 (5.41–77.9)	<0.001
CDAI low disease activity (≤10.0) at wk 12	89 (32.8)	132 (16.2)	2.52 (1.84–3.45)	<0.001
CDAI remission (≤2.8) at wk 24	37 (13.7)	4 (0.4)	35.6 (13.6–129)	<0.001
CDAI low disease activity (≤10.0) at wk 24	119 (43.9)	189 (23.3)	2.58 (1.93–3.45)	<0.001

* CDAI, Clinical Disease Activity Index; CI, confidence interval; LOCF, last observation carried forward; MID, minimal important difference; OR, odds ratio; TNFi, tumor necrosis factor inhibitor.

unmeasured strong predictors of clinical response to new b/tsDMARD initiation were unbalanced between the cohorts (ie, the risk of residual confounding due to RA-related characteristics is low). In addition, we adhered to published guidance for the use of external comparators, including similar geography and treatment history and identical outcome measurement between the prospective cohort and external comparator.[19]

Patients in the external comparator arm were classified as having initiated therapies based upon prescribing data identified in the EHR, similar to an intention-to-treat analysis, and not actual receipt of treatment. It is possible that participants in the PrismRA cohort were more likely to actually receive the chosen study treatment compared to patients in the external comparator cohort, as RA treatment initiation was confirmed for all included participants during the conduct of the prospective study. Of additional note, adherence to study treatment during follow-up (ie, early discontinuation or switching) was not considered in the analysis of clinical outcomes in either cohort because we wished to assess the beneficial effect of PrismRA testing irrespective of subsequent treatment changes, as may occur in real-world practice.

This study had additional limitations that need to be considered in interpreting its findings. Although the comparator patient selection was performed with the most recent b/tsDMARD start dates possible, approximately 33% of the comparator patients initiated therapy during 2018 through 2021, and they may have received different clinical care compared to the PrismRA cohort due to the COVID-19 pandemic or other factors. Given the requirements for the CDAI availability in prespecified time windows, the primary outcome was missing in 25.5% of PrismRA and 41.5% of external comparator patients. Compared to a complete case analysis, LOCF imputation produced worse clinical outcomes for the PrismRA cohort versus the external comparator cohort (ie, LOCF biased results toward the null hypothesis). Outcome assessments in the PrismRA cohort were performed by the treating physician (ie, not blinded) and may have been influenced by PrismRA MSRC results or b/tsDMARD selection.

In conclusion, the use of PrismRA MSRC results to inform selection of TNFi versus non-TNFi b/tsDMARDs in patients with moderate-high RA disease activity was associated with better clinical outcomes compared to matched external comparator patients who did not receive PrismRA testing. PrismRA testing helps fill the need for a precision medicine approach to more rapidly identify the most effective therapy for individual patients with RA.

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 Curtis 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

Scipher Medicine collected the data for the PrismRA prospective cohort. Scipher Medicine had no role in the study design, the collection of data for the external comparator arm, the 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 Scipher Medicine.

REFERENCES

1. 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.
2. Smolen JS, Landewé RBM, Bergstra SA, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. Ann Rheum Dis 2023;82(1):3–18.
3. Jin Y, Desai RJ, Liu J, et al. Factors associated with initial or subsequent choice of biologic disease-modifying antirheumatic drugs for treatment of rheumatoid arthritis. Arthritis Res Ther 2017;19(1):159.
4. Curtis JR, Jain A, Askling J, et al. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. Semin Arthritis Rheum 2010;40(1):2–14.e1.
5. Hazlewood GS, Schieir O, Bykerk V, et al. Frequency of symptomatic adverse events in rheumatoid arthritis: an exploratory online survey. J Rheumatol 2022;49(9):998–1005.
6. Meissner B, Trivedi D, You M, et al. Switching of biologic disease modifying anti-rheumatic drugs in patients with rheumatoid arthritis in a real world setting. J Med Econ 2014;17(4):259–265.
7. Sharma SD, Bluett J. Towards personalized medicine in rheumatoid arthritis. Open Access Rheumatol 2024;16:89–114.
8. Pappas DA, Brittle C, Mossell JE III, et al. Perceived clinical utility of a test for predicting inadequate response to TNF inhibitor therapies in rheumatoid arthritis: results from a decision impact study. Rheumatol Int 2021;41(3):585–593.
9. Cohen S, Wells AF, Curtis JR, et al. A molecular signature response classifier to predict inadequate response to tumor necrosis factor-α inhibitors: the NETWORK-004 prospective observational study. Rheumatol Ther 2021;8(3):1159–1176.
10. Felson DT, Anderson JJ, Boers M, et al. American college of rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis & Rheum 1995;38(6):727–735.
11. Jones A, Rapisardo S, Zhang L, et al. Analytical and clinical validation of an RNA sequencing-based assay for quantitative, accurate evaluation of a molecular signature response classifier in rheumatoid arthritis. Expert Rev Mol Diagn 2021;21(11):1235–1243.
12. Strand V, Zhang L, Arnaud A, et al. Improvement in clinical disease activity index when treatment selection is informed by the tumor necrosis factor-α inhibitor molecular signature response classifier: analysis from the study to accelerate information of molecular signatures in rheumatoid arthritis. Expert Opin Biol Ther 2022;22(6):801–807.
13. Beukelman T, Mudano A, Stewart P, et al. Generating real-world evidence from the excellence network in rheumatology. Pharmacoepidemiol Drug Saf 2024;33(12):e70067.
14. Curtis JR, Strand V, Golombek SJ, et al. Decision impact analysis to measure the influence of molecular signature response classifier

testing on treatment selection in rheumatoid arthritis. Rheumatol Ther 2024;11(1):61–77.

15. Curtis JR, Strand V, Golombek S, et al. Patient outcomes improve when a molecular signature test guides treatment decision-making in rheumatoid arthritis. Expert Rev Mol Diagn 2022;22(10):1–10.
16. Strand V, Cohen SB, Curtis JR, et al. Clinical utility of therapy selection informed by predicted nonresponse to tumor necrosis factor-α inhibitors: an analysis from the Study to Accelerate Information of Molecular Signatures (AIMS) in rheumatoid arthritis. Expert Rev Mol Diagn 2022; 22(1):101–109.
17. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569–2581.
18. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23(5) (suppl 39):S100–S108.
19. Thorlund K, Dron L, Park JJH, et al. Synthetic and external controls in clinical trials - a primer for researchers. Clin Epidemiol 2020;12: 457–467.
20. US Food and Drug Administration. Considerations for the design and conduct of externally controlled trials for drug and biological products. February 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-design-and-conduct-externally-controlled-trials-drug-and-biological-products
21. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009;28(25):3083–3107.
22. Michaud K, Wolfe F. Comorbidities in rheumatoid arthritis. Best Pract Res Clin Rheumatol 2007;21(5):885–906.
23. Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Med Decis Making 2009;29(6): 661–677.
24. Normand ST, Landrum MB, Guadagnoli E, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol 2001;54(4):387–398.
25. Wang SV, Jin Y, Fireman B, et al. Relative performance of propensity score matching strategies for subgroup analyses. Am J Epidemiol 2018;187(8):1799–1807.
26. Williamson E, Morley R, Lucas A, et al. Propensity scores: from naive enthusiasm to intuitive understanding. Stat Methods Med Res 2012; 21(3):273–293.
27. Stuart EA. Matching methods for causal inference: a review and a look forward. Stat Sci 2010;25(1):1–21.
28. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stat 2011;10(2):150–161.
29. Rosenbaum P, Rubin D. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am Stat 1985;39(1):33–38.
30. Austin PC. Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. Am J Epidemiol 2010;172(9):1092–1097.
31. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011;46(3):399–424.
32. Curtis JR, Yang S, Chen L, et al. Determining the minimally important difference in the clinical disease activity index for improvement and worsening in early rheumatoid arthritis patients. Arthritis Care Res (Hoboken) 2015;67(10):1345–1353.
33. Wells G, Li T, Maxwell L, et al. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. J Rheumatol 2007;34(2):280–289.
34. Ward MM, Castrejon I, Bergman MJ, et al. Minimal clinically important improvement of Routine Assessment of Patient Index Data 3 in rheumatoid arthritis. J Rheumatol 2019;46(1):27–30.
35. Salaffi F, Stancati A, Silvestri CA, et al. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004;8(4):283–291.
36. Myles PS, Myles DB, Galagher W, et al. Measuring acute postoperative pain using the visual analog scale: the minimal clinically important difference and patient acceptable symptom state. Br J Anaesth 2017; 118(3):424–429.
37. Curtis JR, Palmer JL, Reed GW, et al. Real-world outcomes associated with methotrexate, sulfasalazine, and hydroxychloroquine triple therapy versus tumor necrosis factor inhibitor/methotrexate combination therapy in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2021;73(8):1114–1124.
38. Broadwell A, Schechtman J, Conaway D, et al. Effectiveness and safety of intravenous golimumab with and without concomitant methotrexate in patients with rheumatoid arthritis in the prospective, noninterventional AWARE study. BMC Rheumatol 2023;7(1):5.
39. Pappas DA, St John G, Etzel CJ, et al. Comparative effectiveness of first-line tumour necrosis factor inhibitor versus non-tumour necrosis factor inhibitor biologics and targeted synthetic agents in patients with rheumatoid arthritis: results from a large US registry study. Ann Rheum Dis 2021;80(1):96–102.

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1159–1166
DOI 10.1002/acr.80009
© 2026 American College of Rheumatology

American College of Rheumatology
Empowering Rheumatology Professionals

Impact of Dual Rheumatoid Factor and Anticitrullinated Protein Antibody Seropositive, Single Seropositive, and Seronegative Rheumatoid Arthritis on Outcomes

Rebecca T. Brooks, MD, Sara J. Achenbach, MS, Vanessa L. Kronzer, MD, MSCI, Elena Myasoedova, MD, PhD, Cynthia S. Crowson, PhD, and John M. Davis III MD, MS

Objective. The objective of this study was to investigate the association between dual seropositive, single seropositive, and seronegative rheumatoid arthritis (RA) with radiographic erosions, disease flares, and death.

Methods. We performed a retrospective, population-based study of residents in Southern Minnesota with incident RA who fulfilled criteria for RA in 2003 to 2019. Radiographic erosions and flares were evaluated within one year of incident RA. All-cause mortality was obtained from medical records and death certificates. Cox models adjusted for age, sex, smoking status, year of incident RA, and comorbidities were used.

Results. The study included 1,373 patients with RA. At RA incidence, 37% were dual seropositive, 13% seropositive for anti–cyclic citrullinated protein (anti-CCP) only, 12% seropositive for rheumatoid factor (RF) only, and 38% seronegative. The highest proportion of radiographic erosions before or within one year of RA incidence was in the dual seropositive (31%) and lowest in those seropositive for anti-CCP only (13%). Flares occurred in 69% of the dual seropositive and 51% of the seropositive for anti-CCP only within the first year of RA incidence. Those seropositive for RF only had over a two-fold increase in mortality rates compared with the seronegatives (adjusted hazard ratio [aHR] 2.18, 95% confidence interval [CI] 1.47–3.24). In contrast, the dual seropositives had only a >50% increase in mortality rates (aHR 1.66, 95% CI 1.21–2.28), and those seropositive only for anti-CCP had a >30% increase in mortality rates (aHR 1.32, 95% CI 0.83–2.11).

Conclusion. Patients with RA seropositive for RF only have an increase in mortality rates compared with patients who are dual seronegative. RF positivity could be an indicator of inflammation that requires pharmacologic treatment to decrease mortality rates.

INTRODUCTION

Rheumatoid arthritis (RA) is a systemic inflammatory disease that predominantly manifests in the joints, leading to swelling, tenderness, and destruction.[1] Patients with RA have higher mortality rates compared with the general population.[2] The most common causes of death in the RA population are cardiovascular disease, cancer, and respiratory diseases.[2,3] Radiographs are frequently used for diagnosis to assess for structural damage in joints at baseline and monitored during treatment to assess for progression of disease.[4] The rates of RA disease flares differ and can be impacted by autoantibody status and the tapering or discontinuation of disease-modifying antirheumatic drugs (DMARDs).[5]

Antibody levels of both rheumatoid factor (RF) and anti–cyclic citrullinated protein (anti-CCP) are included in the 2010 American College of Rheumatology (ACR)/EULAR classification criteria for RA.[1] Previous studies have identified that RF and anti-CCP are associated with increased disease activity and radiographic progression.[6–8] Furthermore, a study using data from the Norfolk Arthritis Register and Leiden Early Arthritis Clinic found patients

Mayo Clinic, Rochester, Minnesota.

The content of this article is solely the responsibility of the authors and does not represent the official views of the NIH or the Mayo Clinic.

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grants R01-AR-46849 and R01-AR-084831); National Institute on Aging (NIA), NIH (grants R01-AG-068192 and K24-AG-078179); the Rheumatology Research Foundation Scientist Development Award (grant FP00119644) (to Dr Kronzer); and the Mark E. and Mary A. Davis Initiative in Rheumatoid Arthritis Research (to Drs Kronzer, Myasoedova, and Davis). This study used the resources of the Rochester Epidemiology Project (REP) medical records–linkage system, which is supported by the NIA (grant AG-058738), by the Mayo Clinic Research Committee, and by fees paid annually by REP users.

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.80009).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80009.

Address correspondence via email to John M. Davis, III, MD, MS, at davis.john4@mayo.edu.

Submitted for publication July 8, 2025; accepted in revised form January 6, 2026.

SIGNIFICANCE & INNOVATIONS

- Previous studies found that rheumatoid factor (RF) and anti–cyclic citrullinated protein (anti-CCP) antibodies are associated with rheumatoid arthritis (RA) disease activity, radiographic progression, and death. However, few studies have investigated how seropositivity for RF and anti-CCP individually influence outcomes.
- Using the Rochester Epidemiology Project, we found that patients seropositive for only RF have over a two-fold increase in mortality rates compared with dual-seronegative patients and that seropositivity for anti-CCP does not further increase mortality rates.
- Given the increased mortality rates found in patients only seropositive for RF, further studies investigating how other factors, such as less commonly identified autoantibodies in RA interact with RF, are necessary to understand the driving mechanism behind the increased mortality rates in patients with RA.

who were dual seropositive for anti-CCP and RF had increased mortality rates compared with subjects who were single seropositive or seronegative.[9] However, the study did not adjust for comorbidities; thus, the increased mortality rates found could be confounded by comorbidities that increase mortality rates.

Few studies have investigated how RF and anti-CCP individually contribute to outcomes in RA such as radiographic changes, disease flares, and mortality rates. Understanding the individual contributions of these serological markers and their impact on disease outcomes will aid clinicians in interpreting these disease markers at diagnosis and the influence on outcomes. The objective of this study was to investigate the association between dual seropositive, single seropositive, and seronegative RA and outcomes. We hypothesized that those seropositive for both RF and anti-CCP would have increased radiographic erosions, disease flares, and mortality rates compared with those who were seronegative.

MATERIALS AND METHODS

Study design. We conducted a retrospective, population-based cohort study of patients with incident RA within the Rochester Epidemiology Project (REP). This large record linkage system provides the medical records of outpatient offices, urgent care offices, emergency departments, and hospitals within Southern Minnesota and Western Wisconsin. Previous studies have demonstrated that the demographics of patients within the REP are like that of the Upper Midwest United States.[10] The counties included were Olmsted, Dodge, Mower, Goodhue, Wabasha, Freeborn, Steele, and Waseca. Patients who met RA criteria in 2003 to 2019 were followed until last medical visit, emigration out of the geographic area, death, or December 31, 2023. Those who did not die were censored at last follow-up.

This study received Mayo Clinic and REP Research Review Committee approval (Institutional Review Board 24-003680 and 006-RRC-24, respectively).

Patient selection. Patients with inflammatory arthritis codes or RF and/or anti-CCP testing were screened by trained nurse abstractors who reviewed the medical records of each potential subject.[11] Individuals with RA were included in the study if the 1987 ACR classification criteria[12] or the 2010 ACR/EULAR classification criteria for RA was fulfilled[1] and the subject was ≥18 years old and resided in the region of interest on the date of incident RA. The RA incidence date was defined as the earliest date of criteria fulfillment. Subjects were excluded if RF and anti-CCP tests were not available before or within one year following the date of incident RA.

Study outcome. The primary outcome in this study was all-cause mortality as determined through electronic records. The REP tracks vital status using medical records and electronic death certificate data from the state of Minnesota and National Death Index Plus. The causes of death were also provided by these sources and were categorized according to *International Classification of Diseases, Ninth and Tenth Versions*. Secondary outcomes included the presence of radiographic erosions before or within one year of RA incidence as documented by either a radiologist or rheumatologist within the electronic medical record (EMR) and a visit for a flare or remission within one year after RA incidence. An RA disease flare was defined as the EMR having documentation of the terms "flare," "active," "ongoing," or "relapse" in relationship to a patient's RA. Additionally, if the EMR had documentation of increasing disease activity, a change of DMARD, or increased use of prednisone due to increased disease activity, the patient was included as having an RA disease flare.[13] If the EMR contained the terms "remission," "quiescent," "inactive," quiet," "asymptomatic," or "no evidence of disease activity" in relation to a patient's RA, the patient was defined as having a visit for RA disease remission.[13]

Study covariates and descriptive variables. The demographics, health behaviors, RA disease characteristics, and use of DMARDs were abstracted by trained nurse abstractors. Race was categorized as non-Hispanic White and non-White. Non-White included patients who identified as Black, Asian, Hawaiian and/or Pacific Islander, American Indian, Hispanic, or other and/or mixed. The Charlson Comorbidity Index was calculated based on the Deyo adaptation. This included 16 comorbidity categories with the exclusion of rheumatic diseases.[14]

The RA disease characteristics included the first documented joint swelling by a physician or advanced practice

provider, baseline erythrocyte sedimentation rate, the presence of rheumatoid nodules, large joint swelling, and extra-articular manifestations. Extra-articular manifestations included neuropathy, serositis, vasculitis, episcleritis, retinal vasculitis, and glomerulonephritis.

Statistical analysis. Chi-square and rank-sum tests were used to compare the baseline characteristics between the cohorts. Cox proportional hazard models adjusted for age, sex, year of RA incidence, smoking status, and Charlson Comorbidity Index were used to determine the association between seropositivity status (dual seropositive, only anti-CCP seropositive, only RF seropositive, and seronegative) and all-cause mortality. Additional models included (1) just age, sex, and year of RA incidence; (2) the original adjustors plus body mass index (BMI) and higher education; and (3) the original adjustors plus DMARDs used in the first year of RA incidence. The model adjusting for DMARDs used in the first year used data starting one year after incidence to avoid immortal time bias. Sensitivity analyses included (1) considering time-dependent RF and anti-CCP status, (2) subdividing positive RF and anti-CCP into low positives (between the upper limit of normal and three times the upper limit of normal) and high positive (>3 times the upper limit of normal), and (3) examining the effect of anti-CCP among the subset of patients who were seropositive for RF. A P value of <0.05 was considered statistically significant for all analyses. Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC) and R version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Baseline patient characteristics. Of the 1,490 patients with incident RA, 1,373 had both RF and anti-CCP tests documented and were included in this study. The patients had a mean age of 55.9 (SD 15.2) years, 69% were women, over half were ever smokers (53%), and 91% were non-Hispanic White. At baseline, 37% were dual seropositive, 13% were seropositive for only anti-CCP, 12% were seropositive for RF only, and 38% were seronegative. More patients in the seronegative cohort (28%) experienced six weeks or more between the first documented joint swelling and RA incidence than those seropositive for RF only (18%) or those dual seropositive or seropositive for only anti-CCP (12% for both; $P < 0.001$). Over half (67%) of all patients with RA were prescribed methotrexate before or within the first year of incident RA, with 78% of the dual seropositive cohort receiving a prescription. Patients who are seropositive only for RF received significantly less prescriptions for methotrexate (57%) (Table 1).

Rheumatoid nodules were seen within the first year after RA incidence in 12% of the dual seropositive cohort, 8% of the RF seropositive–only cohort, 5% of the anti-CCP seropositive–only cohort, and 3% of the seronegative cohort. Interestingly, the highest percentage of extra-articular manifestations within the first year after RA incidence was in the cohort seropositive for RF only (9%), with the lowest percentage (4%) being in those seropositive for only anti-CCP (Table 1).

Association of seropositivity status with radiographic erosions. Over 24% of all patients with RA had radiographic erosions before or within one year of incident RA, with the highest percentage in those who were dual seropositive (31%) and lowest in those seropositive for only anti-CCP (13%). Over a fifth (24%) of patients who were seropositive for only RF had radiographic erosions ($P < 0.001$) (Table 2). Compared with seronegatives, those who were only anti-CCP positive had a lower likelihood of erosions during the first year (adjusted odds ratio 0.57, 95% confidence interval [CI] 0.34–0.95), and the dual seropositives had a higher likelihood of erosions (adjusted odds ratio 1.66, 95% CI 1.23–2.25). Subdividing the groups into those with high and low positive titers revealed the lowest erosion rate (4 of 68; 6%) among patients with negative RF and low positive anti-CCP compared with 17 of 94 (18%) in those with negative RF and high positive anti-CCP. Similarly, among patients with high positive RF, those with low positive anti-CCP had a lower erosion rate than those with high positive anti-CCP (3 of 22, 14% vs 105 of 318, 33%), and among patients with low positive RF, those with low positive anti-CCP had a lower erosion rate than those with high positive anti-CCP (3 of 13, 23% vs 35 of 119, 29%).

Association seropositivity status and RA flares or remission. Visits for disease flares within the first year following incident RA were seen in 62% of all patients with RA, with the highest proportion occurring in the dual seropositive cohort (69%). Compared with seronegatives, those who were only anti-CCP positive had a lower likelihood of flare during the first year (adjusted odds ratio 0.71, 95% CI 0.50–1.00), and the dual seropositives had a higher likelihood of flare (adjusted odds ratio 1.51, 95% CI 1.16–1.95). Nearly a quarter (22%) of the cohort had a visit with documentation of RA remission during the first year following RA criteria. However, there were no significant differences in remission rates between groups (Table 2).

Association of seropositivity with death. During a median of 8.3 (25th to 75th percentile 5.5–11.6) years of follow-up, 237 patients died. When adjusted for age, sex, year of RA incidence, smoking status, and comorbidities, patients who were seropositive for RF only had the largest decrease in survival compared with those who were dual seronegative, with over a two-fold increase in mortality rates (adjusted hazard ratio [aHR] 2.18, 95% CI 1.47–3.24) (Table 3). In the dual seropositives and those only seropositive for anti-CCP, there was an increase in mortality rates compared with the dual seronegatives, but it was less pronounced, with a 66% (aHR 1.66, 95% CI 1.21–2.28) and 32% (aHR 1.32, 95% CI 0.83–2.11) increased risk of death,

Table 1. Characteristics of patients with RA as determined by seropositivity*

Seropositivity status	RF neg/anti-CCP neg (n = 517)	RF neg/anti-CCP pos (n = 179)	RF pos/anti-CCP neg (n = 164)	RF pos/anti-CCP pos (n = 513)	Total (N = 1,373)	*P* value
Demographics						
Age at RA incidence, mean (SD), years	57.0 (15.5)	54.3 (15.5)	56.8 (14.8)	55.1 (14.8)	55.9 (15.2)	0.13
Female sex, n (%)	353 (68)	129 (72)	122 (74)	338 (66)	942 (69)	0.15
Non-Hispanic White, n (%)	491 (95)	161 (90)	151 (93)	446 (87)	1,249 (91)	<0.001
Smoking status, n (%)						<0.001
Never	268 (52)	91 (51)	84 (51)	204 (40)	647 (47)	
Current	62 (12)	33 (18)	33 (20)	134 (26)	262 (19)	
Former	187 (36)	55 (31)	47 (29)	175 (34)	464 (34)	
BMI, mean (SD), kg/m^2	30.6 (7.2)	30.7 (7.2)	30.9 (8.9)	30.0 (6.6)	30.4 (7.2)	0.73
Higher education, n/N (%)	329/481 (68)	106/161 (66)	100/157 (64)	263/462 (57)	798/1261 (63)	0.003
Charlson Comorbidity Index, mean (SD)[a]	1.4 (2.0)	1.2 (1.9)	1.3 (1.6)	1.3 (2.2)	1.3 (2.0)	0.18
RA disease characteristics						
Days from first documented joint swelling to RA incidence, median (25th, 75th percentile)	0 (0, 63)	0 (0, 0)	0 (0, 14)	0 (0, 0)	0 (0,12)	<0.001
≥6 weeks from first documented joint swelling to RA incidence, n (%)	145 (28)	22 (12)	29 (18)	63 (12)	259 (19)	<0.001
ESR, mean (SD), mm/hr	20.3 (20.4)	22.3 (22.1)	20.7 (19.4)	27.5 (22.3)	23.2 (21.4)	<0.001
Rheumatoid nodules, n (%)[b]	16 (3)	9 (5)	13 (8)	59 (12)	97 (7)	<0.001
Large joint swelling, n (%)[b]	286 (55)	75 (42)	64 (39)	265 (52)	690 (50)	<0.001
Extra-articular manifestations, n (%)[b,c]	41 (8)	8 (4)	14 (9)	27 (5)	90 (7)	0.15
DMARD use, n (%)[b]						
Methotrexate	311 (60)	117 (65)	94 (57)	400 (78)	922 (67)	<0.001
Other conventional DMARDs	314 (61)	76 (42)	89 (54)	222 (43)	701 (51)	<0.001
b/ts DMARDs	63 (12)	18 (10)	17 (10)	74 (14)	172 (13)	0.33

* anti-CCP, anti–cyclic citrullinated protein; BMI, body mass index; b/ts, biologic/targeted synthetic; DMARD, disease-modifying antirheumatic drug; ESR, erythrocyte sedimentation rate; neg, negative; pos, positive; RA, rheumatoid arthritis; RF, rheumatoid factor.
[a] Excluding connective tissue diseases.
[b] Before or within one year of incident RA.
[c] Including neuropathy, serositis, vasculitis, episcleritis, scleritis, retinal vasculitis, and glomerulonephritis.

respectively (Table 3). When analyses were adjusted for BMI and higher education or for the use of DMARDs, there were no changes in the results (Table 3). In a subset analysis of just those seropositive for RF, those who were dual seropositive had a 26% lower risk of death (aHR 0.74, 95% CI 0.50–1.09) compared with those who were only seropositive for RF, but this difference did not reach statistical significance. Furthermore, when seropositives were subdivided into low and high positives, patients with only low positive RF titers had the highest risk of death (aHR 2.59, 95% CI 1.59–4.23) compared with seronegatives (Supplementary Table 1). The three most common causes of death among patients with RA were cardiovascular diseases (33%), malignancy (21%), and respiratory diseases (11%). Those seropositive for RF only or dual seropositive had over a four-fold increase in respiratory-related mortality rates (aHR 4.10, 95% CI 1.04–16.24; aHR 4.52, 95% CI 1.57–13.06, respectively) (Table 4). There were 26 deaths related to respiratory diseases, with 62% due to complications of chronic obstructive pulmonary disease, 15% related to interstitial pulmonary disease, and 4% due to pneumonia. There were no significant differences in circulatory-related or malignancy-related deaths between the cohorts.

During follow-up, a total of 28 patients changed from RF seronegative to RF seropositive, 55 patients changed from RF seropositive to RF seronegative, 10 patients changed from anti-CCP seronegative to anti-CCP seropositive, and 24 patients changed from anti-CCP seropositive to anti-CCP seronegative. Sensitivity analyses assessing anti-CCP and RF seropositivity as time-dependent covariates found the results were similar to the primary analysis (Supplementary Table 2).

DISCUSSION

In this study, we aimed to determine the association of seropositivity and radiographic changes before or within the first year of RA incidence, disease flares or remission within the first year of RA incidence, and death in patients with RA. Although we hypothesized that patients with RA with dual seropositivity would have the highest mortality rates as seen in a previous study,[9] we found that patients only seropositive for RF had the

Table 2. Associations of seropositivity with erosions and visits for flares or remission*

Seropositivity status	Patients	Erosions on radiograph[a], n/N (%)	Flare[b], n (%)	Remission[b], n (%)	Erosions on radiograph[a], odds ratio[c] (95% CI)	Flare[b], odds ratio[c] (95% CI)	Remission[b], odds ratio[c] (95% CI)
RF negative and anti-CCP negative	517	101/467 (22)	306 (59)	116 (22)	Reference	Reference	Reference
RF negative and anti-CCP positive	179	21/162 (13)	92 (51)	34 (19)	0.57 (0.34–0.95)	0.71 (0.50–1.00)	0.85 (0.55–1.31)
RF positive and anti-CCP negative	164	35/147 (24)	100 (61)	32 (20)	1.07 (0.68–1.66)	1.11 (0.77–1.59)	0.89 (0.57–1.39)
RF positive and anti-CCP positive	513	146/472 (31)	352 (69)	126 (25)	1.66 (1.23–2.25)	1.51 (1.16–1.95)	1.22 (0.91–1.64)

* anti-CCP, anti–cyclic citrullinated protein; CI, confidence interval; RF, rheumatoid factor.
[a] Before or within one year of incident rheumatoid arthritis.
[b] At least one visit within one year after incident rheumatoid arthritis.
[c] Adjusted for age, sex, year of rheumatoid arthritis incidence, smoking status, and Charlson Comorbidity Index.

highest mortality rates. The previous study by Humphreys et al did not adjust for comorbidities,[9] suggesting that the increased risk of mortality rates seen in dual-seropositive patients was possibly driven by comorbidities. However, in our study with adjustment for comorbidities, dual-seropositive and RF only–seropositive patients with RA continued to have an increase in mortality rates, suggesting that autoimmunity in RA may be an independent driver of mortality rates. Increased mortality rates seen in those seropositive for RF only was independent of treatment with DMARDs. However, this must be interpreted cautiously, given the confounding by indication often encountered in observational studies, and alternative pharmacoepidemiologic study designs will be needed to evaluate the influence of DMARDs. Those with dual seropositivity and seropositivity for RF only had over a four-fold increased risk of respiratory-related death, independent of smoking. This is consistent with previous studies finding an association between seropositivity and elevated risk of respiratory-related death.[15–17] Although patients with RA-related interstitial lung disease are at an elevated risk of death,[18] only a few of the respiratory-related deaths were due to interstitial lung disease in our study. A proposed mechanism for the increased risk of respiratory-related death in RA is worsening of airway obstruction in chronic obstructive pulmonary disease from RA-related inflammation, perhaps due to the lungs being the source of RA-related autoimmunity.[17,19]

It is well established that RF and anti-CCP seropositivity are associated with the development of RA.[20] Anti-CCP seropositivity is an important prognostic factor in RA, and uniquely, RF in patients without RA has been associated with increased mortality rates.[1,21,22] Additionally, RF and anti-CCP have been found to interact in vitro to increase the inflammatory response through RF, increasing the proinflammatory cytokines released by immune complexes for activation of macrophages.[23–25] However, our study suggests that RF may have additional effects on mortality rates from a population level.

A previous study found that the presence of dual seropositivity was associated with increased erosive disease and suggested that RF may increase bone destruction in patients who are also anti-CCP positive.[26] Furthermore, in a prospective study of patients with early RA, those with higher serum levels of anti-CCP and RF were more likely to have radiographic progression in the first two years of follow-up.[27] In our study, we also found an association between dual seropositivity and erosive disease at diagnosis or within one year. There was no association with erosions in the cohort with only seropositivity for RF and decreased likelihood of erosions in patients seropositive for only anti-CCP. However, the CIs were wide, suggesting we would need a larger cohort to detect a statistically significant difference. A compounding effect of anti-CCP and RF on joint inflammation is suggested by significantly more visits for a RA disease flare

Table 3. Associations between seropositivity and all-cause mortality*

Seropositivity status	N patients/N deaths	Adjusted for age, sex, and year of RA incidence, hazard ratio (95% CI)	Adjusted[a] hazard ratio (95% CI)	Adjusted[a] plus DMARDs, hazard ratio (95% CI)	Adjusted[a] plus BMI and higher education, hazard ratio (95% CI)
RF negative and anti-CCP negative	517/77	Reference	Reference	Reference	Reference
RF negative and anti-CCP positive	179/24	1.22 (0.77–1.94)	1.32 (0.83–2.11)	1.40 (0.86–2.26)	1.44 (0.89–2.31)
RF positive and anti-CCP negative	164/41	1.96 (1.33–2.90)	2.18 (1.47–3.24)	2.21 (1.46–3.34)	2.18 (1.44–3.30)
RF positive and anti-CCP positive	513/95	1.81 (1.33–2.46)	1.66 (1.21–2.28)	1.78 (1.26–2.51)	1.69 (1.22–2.36)

* anti-CCP, anti–cyclic citrullinated protein; BMI, body mass index; CI, confidence interval; DMARD, disease-modifying antirheumatic drug; RA, rheumatoid arthritis; RF, rheumatoid factor.
[a] Model adjusted for age, sex, year of RA incidence, smoking status, and Charlson Comorbidity Index.

Table 4. Associations between seropositivity and mortality*

Seropositivity status	n patients / n deaths	All-cause, adjusted hazard ratio[a] (95% CI)	Circulatory, adjusted hazard ratio[a] (95% CI)	Respiratory, adjusted hazard ratio[a] (95% CI)	Neoplasms, adjusted hazard ratio[a] (95% CI)	All-cause, adjusted 10-year mortality rate[a], %	Circulatory, adjusted 10-year mortality rate[a], %	Respiratory, adjusted 10-year mortality rate[a], %	Neoplasms, adjusted 10-year mortality rate[a], %
RF negative and anti-CCP negative	517/77	Reference	Reference	Reference	Reference	13.2	4.8	1.0	3.2
RF negative and anti-CCP positive	179/24	1.32 (0.83–2.11)	0.94 (0.38–2.30)	0.94 (0.11–8.29)	0.67 (0.19–2.33)	16.6	4.2	0.9	2.1
RF positive and anti-CCP negative	164/41	2.18 (1.47–3.24)	1.64 (0.79–3.40)	4.10 (1.04–16.24)	1.72 (0.71–4.17)	22.7	5.9	3.0	4.5
RF positive and anti-CCP positive	513/95	1.66 (1.21–2.28)	1.49 (0.85–2.58)	4.52 (1.57–13.06)	1.28 (0.64–2.52)	18.1	5.9	3.7	3.6

* anti-CCP, anti–cyclic citrullinated protein; CI, confidence interval; RF, rheumatoid factor.
[a] Model adjusted for age, sex, year of rheumatoid arthritis incidence, smoking status, and Charlson Comorbidity Index.

within one year of diagnosis among the dual seropositives as compared with seronegatives. However, those seropositive for only CCP had the lowest flare rate. This could be partially explained by remission or low disease activity only occurring in 40% to 50% of patients, and given lack of response to treatment, medication changes occur that could explain the higher percentage of RA disease flares seen in the year following RA incidence in our study.[5,28]

Although seropositive RA is described as having detectable RF and anti-CCP, other anticitrullinated, anticarbamylated (anti-Carb), antiacetylated (anti-KAc), and malondialdehyde acetaldehyde modified protein antibodies have been associated with RA.[29–33] A recent study demonstrated that even in the absence of anti-CCP antibodies, antimodified protein antibodies (AMPA) are associated with RA.[34] A separate study from Sweden demonstrated that 30% of patients who were anti-CCP seronegative were positive for other anticitrullinated protein antibodies (ACPA) fine specificities and 16% were seropositive for anti-Carb antibodies.[35] Furthermore, one study reported that ACPAs demonstrate multireactivity to posttranslational modifications, including citrullination, acetylation, and carbamylation, and suggests that anti-CCP–seronegative patients may still have other autoantibodies, such as ACPA fine specificities, anti-KAc, or anti-Carb.[36] These studies demonstrate that patients with RA who are classically defined as seronegative may have other AMPA that contribute to the development and progression of their RA.

Limitations to this study include the predominantly non-Hispanic White population, which could limit generalizability. Additional limitations include the small sample size within groups, potential for misclassification of seropositivity status, and variations of clinical assays for RF and anti-CCP. Radiographic erosions and visits for RA disease flares were only assessed during the first year of RA diagnosis, which limits our ability to make conclusions about long-term disease activity. The presence of radiographic erosions was dependent on rheumatologist or radiologist identification and, given that there are not validated definitions for an RA flare or remission, there could have been misclassification of the secondary outcomes. Pharmacoepidemiologic studies investigating treatment patterns are necessary to better understand the possible drivers of the differences in mortality rates based on seropositivity, and further studies investigating the association between other ACPA fine specificities and AMPA on radiographic changes, disease flares, and mortality rates are needed to understand the impact of autoantibody status in RA.

In summary, patients with RA seropositive for RF only have a significant increase in mortality rates compared with dual-seronegative patients. Anti-CCP positivity did not result in a compounded increase in mortality rates in this study. However, dual seropositivity was associated with a slight increase in RA disease flares. Overall, seropositivity for RF may indicate a high risk for death that warrants early initiation of DMARD therapy.

REFERENCES

1. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569–2581. PubMed doi:https://doi.org/10.1002/art.27584
2. Sokka T, Abelson B, Pincus T. Mortality in rheumatoid arthritis: 2008 update. Clin Exp Rheumatol 2008;26(5 Suppl 51):S35–S61. PubMed
3. England BR, Sayles H, Michaud K, et al. Cause-specific mortality in male US veterans with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2016;68(1):36–45. PubMed doi:https://doi.org/10.1002/acr.22642
4. Park YJ, Gherghe AM, van der Heijde D. Radiographic progression in clinical trials in rheumatoid arthritis: a systemic literature review of trials performed by industry. RMD Open 2020;6(2):e001277. PubMed doi:https://doi.org/10.1136/rmdopen-2020-001277
5. Gravallese EM, Firestein GS. Rheumatoid arthritis - common origins, divergent mechanisms. N Engl J Med 2023;388(6):529–542. PubMed doi:https://doi.org/10.1056/NEJMra2103726
6. De Rycke L, Peene I, Hoffman IE, et al. Rheumatoid factor and anticitrullinated protein antibodies in rheumatoid arthritis: diagnostic value, associations with radiological progression rate, and extra-articular manifestations. Ann Rheum Dis 2004;63(12):1587–1593. PubMed doi:https://doi.org/10.1136/ard.2003.017574
7. Quinn MA, Gough AK, Green MJ, et al. Anti-CCP antibodies measured at disease onset help identify seronegative rheumatoid arthritis and predict radiological and functional outcome. Rheumatology (Oxford) 2006;45(4):478–480. PubMed doi:https://doi.org/10.1093/rheumatology/kei203
8. Rönnelid J, Wick MC, Lampa J, et al. Longitudinal analysis of citrullinated protein/peptide antibodies (anti-CP) during 5 year follow up in early rheumatoid arthritis: anti-CP status predicts worse disease activity and greater radiological progression. Ann Rheum Dis 2005;64(12):1744–1749. PubMed doi:https://doi.org/10.1136/ard.2004.033571
9. Humphreys JH, van Nies JA, Chipping J, et al. Rheumatoid factor and anti-citrullinated protein antibody positivity, but not level, are associated with increased mortality in patients with rheumatoid arthritis: results from two large independent cohorts. Arthritis Res Ther 2014;16(6):483. PubMed doi:https://doi.org/10.1186/s13075-014-0483-3
10. St Sauver JL, Grossardt BR, Yawn BP, et al. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. Am J Epidemiol 2011;173(9):1059–1068. PubMed doi:https://doi.org/10.1093/aje/kwq482
11. Myasoedova E, Davis J, Matteson EL, et al. Is the epidemiology of rheumatoid arthritis changing? Results from a population-based incidence study, 1985-2014. Ann Rheum Dis 2020;79(4):440–444. PubMed doi:https://doi.org/10.1136/annrheumdis-2019-216694
12. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988;31(3):315–324. PubMed doi:https://doi.org/10.1002/art.1780310302
13. Raheel S, Matteson EL, Crowson CS, et al. Improved flare and remission pattern in rheumatoid arthritis over recent decades: a population-based study. Rheumatology (Oxford) 2017;56(12):2154–2161. PubMed doi:https://doi.org/10.1093/rheumatology/kex352
14. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45(6):613–619. PubMed doi:https://doi.org/10.1016/0895-4356(92)90133-8
15. Joerns EK, Lopez-Ruiz A, Lennon RJ, et al. Differences in cause-specific mortality in patients with rheumatoid arthritis by sex and seropositivity. J Rheumatol 2025;52(12):1238–1244. PubMed doi:https://doi.org/10.3899/jrheum.2025-0361
16. Johnson TM, Yang Y, Roul P, et al. A narrowing mortality gap: temporal trends of cause-specific mortality in a national matched cohort study in US veterans with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2023;75(8):1648–1658. PubMed doi:https://doi.org/10.1002/acr.25053
17. Sparks JA, Chang SC, Liao KP, et al. Rheumatoid arthritis and mortality among women during 36 years of prospective follow-up: results from the Nurses' Health Study. Arthritis Care Res (Hoboken) 2016;68(6):753–762. PubMed doi:https://doi.org/10.1002/acr.22752
18. Samhouri BF, Vassallo R, Achenbach SJ, et al. Incidence, risk factors, and mortality of clinical and subclinical rheumatoid arthritis-associated interstitial lung disease: a population-based cohort. Arthritis Care Res (Hoboken) 2022;74(12):2042–2049. PubMed doi:https://doi.org/10.1002/acr.24856
19. Demoruelle MK, Weisman MH, Simonian PL, et al. Brief report: airways abnormalities and rheumatoid arthritis-related autoantibodies in subjects without arthritis: early injury or initiating site of autoimmunity? Arthritis Rheum 2012;64(6):1756–1761. PubMed doi:https://doi.org/10.1002/art.34344
20. Bizzaro N, Bartoloni E, Morozzi G, et al; Forum Interdisciplinare per la Ricerca nelle Malattie Autoimmuni (FIRMA Group). Anti-cyclic citrullinated peptide antibody titer predicts time to rheumatoid arthritis onset in patients with undifferentiated arthritis: results from a 2-year prospective study. Arthritis Res Ther 2013;15(1):R16. PubMed doi:10.1186/ar4148
21. Liang KP, Kremers HM, Crowson CS, et al. Autoantibodies and the risk of cardiovascular events. J Rheumatol 2009;36(11):2462–2469. PubMed doi:https://doi.org/10.3899/jrheum.090188
22. Tomasson G, Aspelund T, Jonsson T, et al. Effect of rheumatoid factor on mortality and coronary heart disease. Ann Rheum Dis 2010;69(9):1649–1654. PubMed doi:https://doi.org/10.1136/ard.2009.110536
23. Anquetil F, Clavel C, Offer G, et al. IgM and IgA rheumatoid factors purified from rheumatoid arthritis sera boost the Fc receptor- and complement-dependent effector functions of the disease-specific anti-citrullinated protein autoantibodies. J Immunol 2015;194(8):3664–3674. PubMed doi:https://doi.org/10.4049/jimmunol.1402334
24. Laurent L, Anquetil F, Clavel C, et al. IgM rheumatoid factor amplifies the inflammatory response of macrophages induced by the rheumatoid arthritis-specific immune complexes containing anticitrullinated protein antibodies. Ann Rheum Dis 2015;74(7):1425–1431. PubMed doi:https://doi.org/10.1136/annrheumdis-2013-204543
25. Sokolove J, Johnson DS, Lahey LJ, et al. Rheumatoid factor as a potentiator of anti-citrullinated protein antibody-mediated inflammation in rheumatoid arthritis. Arthritis Rheumatol 2014;66(4):813–821. PubMed doi:https://doi.org/10.1002/art.38307
26. Hecht C, Englbrecht M, Rech J, et al. Additive effect of anticitrullinated protein antibodies and rheumatoid factor on bone erosions in patients with RA. Ann Rheum Dis 2015;74(12):2151–2156. PubMed doi:https://doi.org/10.1136/annrheumdis-2014-205428
27. Vencovský J, Machácek S, Sedová L, et al. Autoantibodies can be prognostic markers of an erosive disease in early rheumatoid arthritis. Ann Rheum Dis 2003;62(5):427–430. PubMed doi:https://doi.org/10.1136/ard.62.5.427
28. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. JAMA 2018;320(13):1360–1372. PubMed doi:https://doi.org/10.1001/jama.2018.13103
29. Ahn JK, Hwang J, Chang Y, et al. Rheumatoid factor positivity increases all-cause and cancer mortality: a cohort study. Rheumatol Int 2017;37(7):1135–1143. PubMed doi:https://doi.org/10.1007/s00296-017-3738-x
30. Jiang X, Trouw LA, van Wesemael TJ, et al. Anti-CarP antibodies in two large cohorts of patients with rheumatoid arthritis and their relationship to genetic risk factors, cigarette smoking and other

autoantibodies. Ann Rheum Dis 2014;73(10):1761–1768. PubMed doi:https://doi.org/10.1136/annrheumdis-2013-205109

31. Juarez M, Bang H, Hammar F, et al. Identification of novel antiacetylated vimentin antibodies in patients with early inflammatory arthritis. Ann Rheum Dis 2016;75(6):1099–1107. PubMed doi:https://doi.org/10.1136/annrheumdis-2014-206785
32. Shi J, Knevel R, Suwannalai P, et al. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. Proc Natl Acad Sci U S A 2011;108(42):17372–17377. PubMed doi:https://doi.org/10.1073/pnas.1114465108
33. Thiele GM, Duryee MJ, Anderson DR, et al. Malondialdehyde-acetaldehyde adducts and anti-malondialdehyde-acetaldehyde antibodies in rheumatoid arthritis. Arthritis Rheumatol 2015;67(3):645–655. PubMed doi:https://doi.org/10.1002/art.38969
34. Grönwall C, Liljefors L, Bang H, et al. A comprehensive evaluation of the relationship between different IgG and IgA anti-modified protein autoantibodies in rheumatoid arthritis. Front Immunol 2021;12:627986. PubMed doi:https://doi.org/10.3389/fimmu.2021.627986
35. Reed E, Hedström AK, Hansson M, et al. Presence of autoantibodies in "seronegative" rheumatoid arthritis associates with classical risk factors and high disease activity. Arthritis Res Ther 2020;22(1):170. PubMed doi:https://doi.org/10.1186/s13075-020-02191-2
36. Sahlström P, Hansson M, Steen J, et al. Different hierarchies of anti-modified protein autoantibody reactivities in rheumatoid arthritis. Arthritis Rheumatol 2020;72(10):1643–1657. PubMed doi:https://doi.org/10.1002/art.41385

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1167–1176
DOI 10.1002/acr.80013
Published 2026. This article is a U.S. Government work and is in the public domain in the USA. *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Evaluating a Pragmatic Strength Alternative for Frailty Measurement and Assessing Its Predictive Capacity Against Established Frailty Instruments in Rheumatoid Arthritis

Kylie E. Riggles, BS,[1] Hannah F. Brubeck, BS,[1] Adrienne D. Tanus, MPH,[1] Courtney N. Loecker, PhD, MSN, APRN-NP,[2,3] Punyasha Roul, MS,[2] Bryant R. England, MD, PhD,[2,3] Elizabeth R. Wahl, MD, MAS,[1,4] James S. Andrews, MD,[5,6,7] Namrata Singh, MD,[4] Joshua F. Baker, MD, MSCE,[8,9] Patricia P. Katz, PhD,[10] Dolores M. Shoback, MD,[10,11] Jose M. Garcia, MD, PhD,[1,4,12] Ariela R. Orkaby, MD, MPH,[13,14,15] and Katherine D. Wysham, MD[1,4,12]

Objective. Frailty occurs prematurely in rheumatoid arthritis (RA) and is associated with poor health outcomes. We compared the performance of four frailty instruments, including a pragmatic alternative measure using chair sit-to-stand (STS), and evaluated their abilities to predict poor health outcomes.

Methods. Frailty was measured at baseline using four instruments: the Fried Frailty Phenotype with STS (Fried-STS), the Fried Frailty Phenotype with hand grip strength (Fried-HGS), the Veterans Affairs Frailty Index (VA-FI), and the FRAIL Scale. Outcomes collected at the one-year follow-up included category of falls (none, one, more than one), category of days hospitalized (none, one to three, more than three), and a composite outcome of fall, hospitalization, or death. Ordinal logistic or logistic regression models, adjusted for age and sex, explored the association of frailty and each outcome.

Results. A total of 143 participants were included (mean ± SD age 64.5 ± 11.7 years, 73% male, and 69% White). Categorization as frail differed by instrument: Fried-STS, 17%; Fried-HGS, 15%; VA-FI, 36%; and FRAIL Scale, 20%. There was poor agreement between frailty instruments (k = 0.07–0.31) except for the Fried-STS and Fried-HGS (k = 0.62). Frailty by the Fried-STS, Fried-HGS, and FRAIL Scale was associated with falls (adjusted odds ratios [aORs] 3.83–9.54, $P < 0.05$). Frailty by the VA-FI was associated with days hospitalized (aOR 5.21, $P = 0.017$). Frailty by the Fried-STS, VA-FI, and FRAIL Scale was associated with higher odds of the composite measure of incident fall, hospitalization, or death (aORs 2.93–7.25, $P < 0.05$).

Conclusion. Each frailty measure predicted adverse health outcomes, with phenotypic and patient-reported measures predicting falls and the deficit accumulation model predicting hospitalization days. Being frail by the Fried-HGS did not predict poor outcomes as well as the other frailty instruments, including the Fried-STS.

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

Supported by the VA Clinical Studies Research and Development (CSR&D) (IK2 CX002351) and the VA Puget Sound Health Care System Geriatrics Research Education and Clinical Center. Dr England's work was supported by the VA CSR&D (CX002203) and the Rheumatology Research Foundation. Dr Singh's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant K23-AR-079588) and the National Institute on Aging, NIH (grant R03-AG-082857). Dr Baker's work was supported by a Rehabilitation Research & Development merit award (RX003644). Dr Garcia's work was supported by the US Department of Veterans Affairs (BX002807), the Congressionally Directed Medical Research Programs (PC170059), and the NIH (grants R01-CA-239208, R01-AG-061558, and R01-CA-279220).

[1]VA Puget Sound Health Care System, Seattle, Washington; [2]University of Nebraska Medical Center, Omaha; [3]VA Nebraska-Western Iowa Health Care System, Omaha; [4]University of Washington School of Medicine, Seattle; [5]University of Alabama at Birmingham; [6]Birmingham VA Health Care System, Birmingham, Alabama; [7]Birmingham/Atlanta VA Geriatrics Research Education and Clinical Center, Birmingham, Alabama; [8]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania; [9]University of Pennsylvania, Philadelphia; [10]University of California, San Francisco; [11]San Francisco VA Medical Center, San Francisco, California; [12]VA Puget Sound Geriatric Research Education and Clinical Center, Seattle, Washington; [13]New England Geriatric Research, Education, and Clinical Center, Boston, Massachusetts; [14]VA Boston Healthcare System, Boston, Massachusetts; [15]Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts.

Authors Riggles and Brubeck 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.80013).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80013.

Address correspondence via email to Katherine D. Wysham, MD, at kwysham@uw.edu.

Submitted for publication June 10, 2025; accepted in revised form January 22, 2026.

SIGNIFICANCE & INNOVATIONS

- The use of chair sit-to-stand as an alternative strength measurement in the Fried Frailty Phenotype showed strong agreement with the traditional hand grip strength measure in our cohort.
- Distinct constructs of frailty demonstrated poor agreement with each other, but each predicted poor health outcomes, with phenotypic and patient-reported measures predicting falls and the deficit accumulation model predicting days spent hospitalized.
- The Fried Frailty Phenotype using chair sit-to-stand demonstrated better prediction of poor health outcomes than the traditional measure with hand grip strength and may offer a pragmatic alternative for frailty measurement in people with rheumatoid arthritis.

INTRODUCTION

Frailty is a state of diminished physiologic reserve that confers increased vulnerability to adverse health outcomes, including falls, fractures, hospitalization, and mortality, in the general population.[1,2] Frailty occurs at a younger age and has higher prevalence in individuals with rheumatoid arthritis (RA) when compared to the general population.[3–6] The burden of frailty in RA increases with age and multimorbidity,[7,8] making promotion of healthy aging critical for adults with RA. Despite the recent advances of frailty research in RA, it is unknown whether frailty tools derived in the general population perform similarly in RA.[3,9]

Most frailty assessments are guided by two frameworks: phenotypic and deficit accumulation models.[10,11] Phenotypic models conceptualize frailty as a physical syndrome reflecting decreased biologic reserve and resilience.[12] In contrast, the deficit accumulation approach views frailty as a cumulative state resulting from the burden of health deficits.[10,13] RA may impact frailty ascertainment regardless of the model used. RA disease activity often results in pain and fatigue,[14] which may confound phenotypic frailty assessment. On the other hand, deficit accumulation methods can be impacted by the accrual of comorbidities directly due to RA and secondary effects from RA treatments.[7,8] Importantly, hand grip strength (HGS), the strength measure used in the Fried Frailty Phenotype to estimate whole-body strength, may not be an accurate measure of whole-body strength in RA due to hand deformities and pain.[15] Additionally, HGS requires a hand dynamometer to obtain readings, which is not widely available in routine clinical care. Chair sit-to-stand (STS) is a valid measure of lower body strength[16] that has been used to measure frailty in the general population[17] and has been validated as an alternative strength measure to measure phenotypic frailty in other disease states[18,19]; however, this has yet to be explored in RA.

We aimed to evaluate the use of STS as a pragmatic alternative to HGS for phenotypic frailty ascertainment by comparing it with, and evaluating its ability to predict poor health outcomes against, three validated frailty instruments: a phenotypic model, the Fried Frailty Phenotype with HGS (Fried-HGS)[1]; a deficit accumulation model, the Veterans Affairs Frailty Index (VA-FI)[20]; and the patient-reported FRAIL Scale.[21] We hypothesized that these frailty measures, which are based on different conceptual frameworks, would demonstrate poor agreement in frailty classification, yet all would equally predict adverse outcomes.

METHODS

Participants. This study analyzed data from the VA Rheumatoid Arthritis Frailty and Osteoporosis (VA-FROst) study, an ongoing single-center, longitudinal cohort study of US Veterans with RA at the VA Puget Sound Health Care System (VAPSHCS). Participants were required to have a clinical RA diagnosis from a board-certified rheumatologist and meet the 2010 American College of Rheumatology/EULAR classification criteria for RA.[22] Eligibility criteria for the primary study focused on osteoporosis, with a minimum age of 35 years to reflect achievement of peak bone mass, and also excluded those with other forms of inflammatory arthritis, medical conditions affecting bone health, or diseases impairing muscle function. Additional exclusions for the parent study are detailed in Supplemental Figure 1. The study was sanctioned by the VAPSHCS Institutional Review Board (IRB no. 1872726), and all participants provided written informed consent.

Study design. All participants completed a baseline in-person visit that included frailty measurement, collection of RA disease characteristics, and other assessments. For this analysis, participants without three years of clinical data within the VA were excluded because it prevented the accurate calculation of the VA-FI (n = 2). At the time of this analysis, all but seven participants had either completed their one-year follow-up visit or had died, allowing for contribution of health outcome data. The remaining participants had either withdrawn from the study before follow-up (n = 5) or had not yet completed their follow-up visit (n = 2) and were therefore excluded from our longitudinal analyses evaluating poor health outcomes (Supplemental Figure 1).

Outcome measures. The outcome measures were number of falls, days spent hospitalized, and a composite of any fall, all-cause nonelective hospitalization, or death. Falls and hospitalizations were recalled over six-month time intervals and self-reported at a follow-up six-month phone call and one-year in-person follow-up visit (Supplemental Figure 1). Falls were categorized as none, one, or multiple (more than one) falls. Number of days hospitalized were confirmed by chart review. Hospitalization days were categorized based on severity of hospitalization as zero days, one to three days, and more than three days.[23] Death occurrence was determined through chart review or by report

from the next-of-kin during follow-up contact attempts by the study team. There is no missingness in the death data due to the nature of the study design requiring consistent follow-up contacts.

Predictors: Frailty measures. *Fried-HGS and Fried-STS.* The Fried-HGS assesses phenotypic frailty by measuring exhaustion, physical activity levels, weight changes, walking speed, and HGS (Supplemental Table 1).[1] Self-reported exhaustion was assessed using the Center for Epidemiologic Studies Depression Scale (CES-D).[24] Participants completed the International Physical Activity Questionnaire to quantify physical activity.[25] Weight loss was defined as >5% over the past year and was ascertained using historical medical record weight (one year before baseline) and weight directly measured in study visits.[1] Fifteen-feet walk speed was measured by one trial.[1] Both HGS and walk speed were assessed at research visits and categorized based on existing cutoffs.[1] HGS was measured by three trials on each hand, and the highest value was used for analysis.[1] Participants earned a point for each subcategory that met the deficit threshold based on their respective cutoffs, and points were summed to assign a score between 0 and 5 for each participant. Based on total deficit score, individuals were categorized as robust (0), prefrail (1–2), or frail (3–5).[1] Chair STS was substituted for HGS when calculating the Fried-STS. Cutoffs for STS were based on a validated cutoff for sarcopenia of >15 seconds to complete five chair stands.[26]

VA-FI. The VA-FI assesses frailty through a deficit accumulation model using over 6,000 administrative claim codes from VA administrative data. These are based on *International Classification of Disease, Ninth Revision* (ICD-9), ICD-10, *Current Procedural Terminology*, and Healthcare Common Procedure Coding System codes and grouped into five domains: morbidities, cognitive and mood disorders, sensory deficits, functional impairments, and other geriatric syndromes[20] (Supplemental Table 2). VA-FI scores were calculated as the ratio of deficits present in an individual out of the total possible deficits (n/31) that were accumulated over a three-year look-back period from the baseline visit date and resulted in a range of 0 to 1. These scores were then placed into frailty categories derived based on previously defined cutoffs: robust (≤0.1), prefrail (>0.1–0.20), and frail (>0.2).[20]

FRAIL Scale. The FRAIL Scale is a patient-reported measure that was constructed to be an overlapping instrument, mixing features of phenotypic and deficit accumulation models. The FRAIL Scale queries fatigue, resistance, ambulation, number of chronic illnesses, and weight loss over the last year[21] (Supplemental Table 3). Points were assigned for each deficit based on existing cutoffs[21] and summed for a total score 0 to 5. Participants were categorized based on their point totals (robust = 0, prefrail = 1–2, frail = 3–5).[21]

Other measures. Sociodemographic data were obtained at baseline through participant self-report and included age, sex, race, ethnicity, and smoking history. Body mass index (BMI) was calculated using height and weight measurements collected during the baseline visit. RA disease duration was derived from the participant-reported date of RA diagnosis. Multidimensional Health Assessment Questionnaire (MDHAQ),[27] pain scores (0–10), and joint replacements data were also obtained by self-report. Anti–cyclic citrullinated peptide positivity was extracted from laboratory records from the electronic medical record. The Disease Activity Score in 28 joints using the erythrocyte sedimentation rate (DAS28-ESR) and DAS28 using c-reactive protein (DAS28-CRP) were collected by a board-certified rheumatologist, participant self-report, and laboratory records and was calculated based on an existing equation on a scale of 0 to 9.4.[28] Presence of deformities was also determined by a board-certified rheumatologist at the baseline visit.

Statistical analysis. The participants' characteristics were described at baseline as well as at the one-year follow-up for the subcohort who contributed follow-up data. Frailty categorization between frailty instruments was visualized by chord diagrams. Frailty instrument components were visualized using pie charts. For the Fried-STS, Fried-HGS, and FRAIL Scale, these were described as n participants (percentage) because each component contributes points equally. For the VA-FI, because each category contains unequal numbers of deficits, these were described as the mean number of deficits in the category (percentage contribution of deficits in the category to the total mean score). Kappa testing was used to assess the agreement between the three frailty levels of the Fried-STS, Fried-HGS, VA-FI, and FRAIL Scale. To ensure that greater BMI did not negatively impact the STS measure, Pearson's correlation test was used to assess whether there was a relationship between STS completion time and BMI.

We explored the associations between three-level frailty measured by instrument (Fried-STS, Fried-HGS, VA-FI, and FRAIL Scale) and poor health outcomes for those who completed the one-year visit. Ordinal logistic regressions were used to evaluate the association between frailty level and the category of falls and category of cumulative days of hospitalization and were reported as odds ratios (ORs). We tested the assumption of the proportion odds using the parallel lines proportion test. Logistic regressions evaluated the association between frailty category and the composite outcome of occurrence of a fall, hospitalization, or death. The robust frailty group was the reference group in all models. Models were performed unadjusted and adjusted for age and sex. Additional covariates were not included in the primary model due to the number of outcomes. Because RA disease activity can influence frailty as well as numerous poor health outcomes, sensitivity analyses were performed to include RA disease activity, measured by DAS28-CRP, as a covariate. Significance level was set at $P < 0.05$, and analyses were performed using STATA version 18.0.[29]

RESULTS

Cohort characteristics. Baseline data were collected from 143 participants with an average age of 64.5 ± 11.7 years (Table 1); 73% were male, 69% were White, 13% were Black, and 12% reported multiple races. Across the survey instruments, the VA-FI classified the highest proportion of veterans as frail (n = 51, 36%), followed by the FRAIL Scale (n = 28, 20%), the Fried-STS (n = 24, 17%), and the Fried-HGS (n = 22, 15%). People who were frail, regardless of the measure used, were older (range of means 67.5–68.5 years), had more pain (range of means 5.0–6.3), and had higher MDHAQ (range of means 0.8–1.1) and DAS28-ESR (range of means 4.0–4.7) scores than the entire cohort. Compared to those classified as frail by the Fried-HGS, participants who were frail by the Fried-STS were less likely to be male (79% vs 92%) and had a shorter disease duration (mean ± SD 13.6 ± 13.2 vs 15.4 ± 15.9 years). In contrast to those classified as frail by the Fried-HGS, those classified as frail by the VA-FI had relatively lower MDHAQ (mean ± SD 0.8 ± 0.4 vs 1.1 ± 0.4), DAS28-ESR (mean ± SD 4.0 ± 1.2 vs 4.6 ± 1.3), and pain scores (mean ± SD 5.0 ± 2.4 vs 6.0 ± 2.6). Individuals classified as frail by the FRAIL Scale had longer disease duration (mean ± SD 17.3 ± 17.7 vs 15.4 ± 15.9 years) than those classified as frail by the Fried-HGS. Those classified as frail by the Fried-HGS had more deformities (n = 9, 41%) than those classified as frail by other measures (Table 1).

Although the Fried-STS and Fried-HGS demonstrated substantial agreement with each other in their frailty categorization of participants (k = 0.62) (Figure 1, Supplemental Table 4), both showed poor agreement when compared to the VA-FI and FRAIL Scale instruments (k = 0.07–0.31) (Figure 2, Supplemental Table 4). Only four (3%) veterans were considered frail by all four measures. There was no correlation between BMI and STS time (r = 0.01) to indicate an impact of BMI on STS performance.

The most frequent deficits in the Fried-STS and Fried-HGS were exhaustion (51%) and low strength; however, more people had low strength using STS (53%) as compared to HGS (39%) (Figure 3A and B). Morbidities (41%) and functional deficits (30%) had the greatest contributions to VA-FI scores, whereas FRAIL Scale components demonstrated similar frequencies (26%–36%) except for five or more illnesses (7%) (Figure 3C and D).

Frailty and adverse health outcomes. One hundred thirty-six individuals had either completed their one-year visit or died, allowing them to contributed follow-up data. Similar demographics and clinical characteristics were observed in the cohort that contributed follow-up data compared to the entire cohort at baseline (Table 2). Sixty-five participants experienced a poor health outcome (fall, hospitalization, or death) and were all included in the composite outcome. Of those, 43 fell but were not hospitalized, 10 were hospitalized but did not fall, and 12 both fell and were hospitalized. Two people died, both of whom also fell and were hospitalized before death.

Table 1. Baseline demographics and clinical characteristics of entire baseline cohort (N = 143) and those determined frail by each measure*

		Participants determined frail by each instrument			
	Total (N = 143)	Fried-STS (n = 24, 17%)	Fried-HGS (n = 22, 15%)	VA-FI (n = 51, 36%)	FRAIL Scale (n = 28, 20%)
Sociodemographic					
Age, mean (SD), y	64.5 (11.7)	67.5 (8.8)	68.2 (8.8)	68.5 (9.6)	67.9 (11.6)
Male sex, n (%)	105 (73)	19 (79)	20 (92)	38 (75)	25 (89)
Race, n (%)					
Black	18 (13)	0 (0)	0 (0)	6 (12)	0 (0)
Multiple races	17 (12)	3 (13)	4 (18)	6 (12)	4 (14)
Other[a]	9 (6)	2 (8)	0 (0)	2 (4)	0 (0)
White	99 (69)	19 (79)	18 (82)	37 (73)	24 (86)
Ethnicity, n (%)					
Hispanic/Latino	9 (6)	0 (0)	0 (0)	0 (0)	2 (7)
BMI, mean (SD)	30.0 (5.5)	28.0 (5.6)	28.5 (5.9)	29.7 (5.3)	31.5 (5.5)
Cigarette smoking, mean (SD), pack years	18.6 (26.8)	30.5 (31.6)	31.6 (33.4)	21.9 (23.1)	23.4 (23.2)
RA disease characteristics					
Disease duration, mean (SD), y	14.6 (13.8)	13.6 (13.2)	15.4 (15.9)	14.3 (13.3)	17.3 (17.7)
MDHAQ score (range 0.0–3.0), mean (SD)	0.7 (0.5)	1.1 (0.4)	1.1 (0.4)	0.8 (0.4)	1.1 (0.4)
Pain score (range 0–10), mean (SD)	4.4 (2.5)	6.0 (2.1)	6.0 (2.6)	5.0 (2.4)	6.3 (2.2)
CCP positive, n (%)	92 (64)	15 (63)	13 (59)	33 (65)	19 (68)
DAS28-ESR (range 0–9.4), mean (SD)	3.7 (1.4)	4.4 (1.1)	4.6 (1.3)	4.0 (1.2)	4.7 (1.4)
Deformities, n (%)	45 (32)	8 (33)	9 (41)	19 (37)	7 (25)
Joint replacements, n (%)	32 (22)	5 (21)	6 (27)	14 (28)	9 (32)

* BMI, body mass index; CCP, cyclic citrullinated peptide; DAS28-ESR, Disease Activity Score in 28 joints using the erythrocyte sedimentation rate; HGS, hand grip strength; MDHAQ, Multidimensional Health Assessment Questionnaire; RA, rheumatoid arthritis; STS, sit-to-stand; VA-FI, Veterans Affairs Frailty Index.

[a] Other race: American Indian, Alaskan Native, Pacific Islander, Asian, other, and unknown.

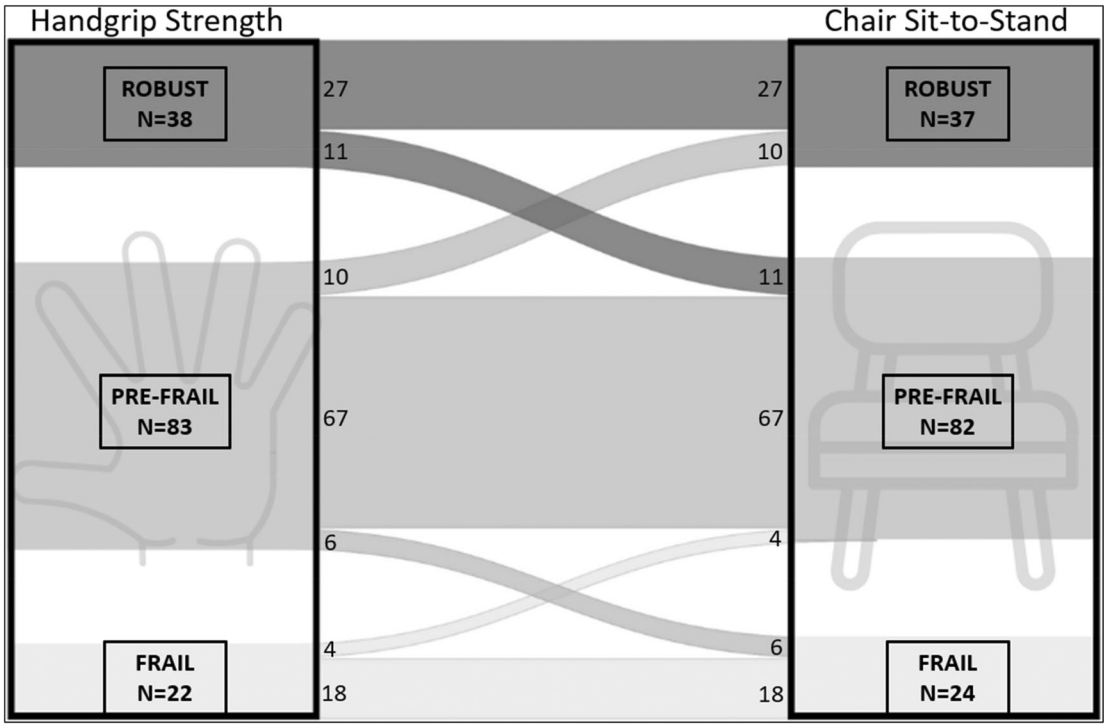

Figure 1. Agreement and recategorization of frailty categories between the traditional Fried Frailty Phenotype with handgrip strength and the substituting chair STS (N = 143). Within-category changes were the most frequent (eg, prefrail → prefrail) and are not pictured in this figure. Within the prefrail category, 21 participants' frailty scores increased with the substitution of STS and six scores decreased. Within the frail category, three scores increased with STS substitution and one score decreased. No changes happened within the robust category, as it requires a total score of 0. STS, sit-to-stand.

Being prefrail by all measures was associated with an increased category of falls (adjusted ORs [aORs] 3.81–4.23, $P < 0.01$), except for the VA-FI (Figure 4A, Supplemental Table 5). Prefrailty was not associated with an increased category of cumulative hospitalization days in any measure (Figure 4B, Supplementary Table 5) but was associated with increased odds of the composite outcome of fall, hospitalization, or death incidence using all frailty measures (aORs 4.41–5.41, $P < 0.01$), except for the VA-FI (Figure 4C, Supplemental Table 5). Being characterized as frail was associated with higher category of falls (aORs 3.83–9.54, $P < 0.05$) in all frailty measures except the VA-FI. The point estimates between the Fried-HGS-measured prefrail and frail groups, however, did not differ (prefrail aOR 3.81 [95% confidence interval (CI) 1.52–9.54]; frail aOR 3.83 [95% CI 1.14–12.94]). However, frailty was only associated with higher category of hospitalization days using the VA-FI and none of the other measures (aOR 5.21, $P = 0.017$). Being frail was also associated with increased odds of the composite incidence outcome in all frailty measures (aORs 2.93–7.25, $P < 0.05$) except for the Fried-HGS (aOR 2.57, $P = 0.118$). The sensitivity analyses including RA disease activity as a covariate demonstrated similar findings to the primary models, with the exception of the Fried-HGS and VA-FI no longer having statistically significant associations with falls and the composite outcome, respectively (Supplemental Table 6).

DISCUSSION

Frailty was common and predicted poor health outcomes in our cohort of US veterans with RA. Consistent with other studies, the VA-FI deficit accumulation model categorized more participants as frail than the other three instruments.[3,5,11] The phenotypic instruments (Fried-STS and Fried-HGS) and the mixed patient-reported FRAIL Scale categorized a similar number of individuals as frail. Although frailty instruments based on different conceptual models had poor agreement, our exploratory analyses found that the Fried-STS and FRAIL Scale similarly predicted falls and the composite outcome, whereas the VA-FI predicted hospitalizations and the composite.[11] Being frail by the traditional Fried-HGS did predict falls; however, the point estimate was similar to that of the prefrail group, unlike the other frailty measures, which showed a stepwise increase between the prefrail and frail groups. The Fried-STS also did not significantly predict days spent hospitalized; however, its ability to predict falls and the composite outcome better than the Fried-HGS suggests that the Fried-STS may be a pragmatic alternative to the Fried-HGS for measuring phenotypic frailty in RA.

Previous studies have validated chair STS as a strength measure in frailty ascertainment in both the general population[17] and some specific disease states, including liver and kidney transplantation allocation,[18,19] but our study is the first to explore this in

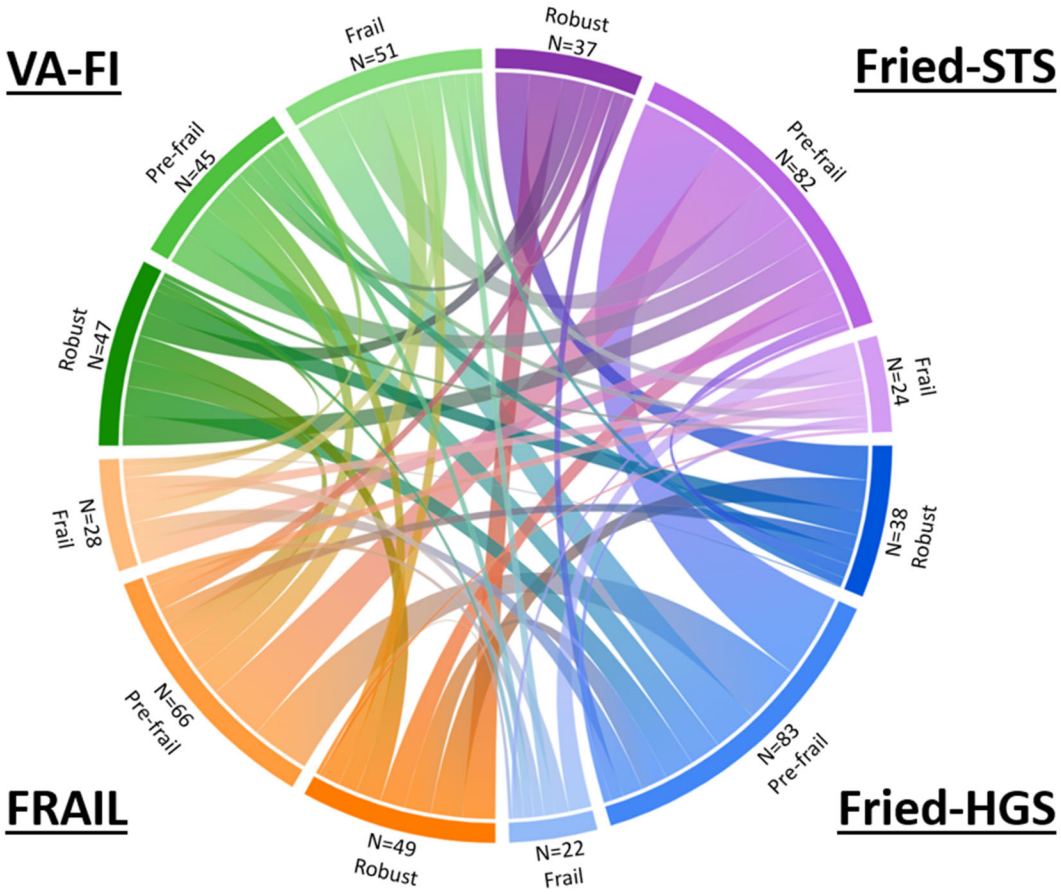

Figure 2. Frailty status by instrument at baseline: Fried-STS, Fried-HGS, VA-FI, and FRAIL Scale. Lines between bars show the agreement or recategorization of frailty categories between measures. Fried-HGS, Fried Frailty Phenotype with handgrip strength; Fried-STS, Fried Frailty Phenotype with chair sit-to-stand; VA-FI, Veterans Affairs Frailty Index.

RA. We found that when using chair STS in place of HGS to measure strength for phenotypic frailty ascertainment, frailty was significantly associated with falls and our composite outcome variable (incidence of fall, hospitalization, or death), aligning with frailty measured by the FRAIL Scale. However, frailty measured by the Fried-HGS only demonstrated a significant association with category of falls in this cohort, and the point estimates between the prefrail and frail groups were comparable, suggesting poor discrimination between frailty levels, unlike the Fried-STS. Although our small sample size may preclude statistically significant associations in those categorized as frail by the Fried-HGS with the composite outcome, the comparably sized subgroups of those identified as frail by the Fried-STS suggest that the Fried-STS may be a more accurate measure of frailty in RA.

It is known that although RA is a multisystemic disease, it often has a disproportionate impact on the joints of the hand, resulting in pain and deformities that can cause altered hand strength and function.[30] Prior work by our group has demonstrated a strong relationship between RA disease activity and frailty, though etiologies of each are difficult to disentangle due to overlapping frailty features such as exhaustion and HGS.[31] In our cohort, those classified as frail using HGS had longer disease duration and more deformities, suggesting HGS may not be an accurate measurement of generalized strength in similar individuals. Our sensitivity analysis revealed that the Fried-HGS lost its only significant association with the poor health outcomes after adjusting for RA disease activity, whereas the Fried-STS's associations, though attenuated, retained their significance, suggesting that the Fried-STS may be less impacted by RA disease activity than the Fried-HGS. Though work has been done to adjust HGS strength-based cutoffs to account for aspects of RA disease,[32] the measurement of HGS may not be feasible in many clinical settings because it requires a specialized hand dynamometer. For this reason, we decided to explore STS as a pragmatic alternative for frailty measurement because it only requires a chair. These findings need to be interpreted in the context of the exploratory nature of our study due to the small number of participants and outcomes available for analysis. We also acknowledge that HGS may be the most appropriate tool to measure strength in individuals with RA who have preserved hand function but have lower-extremity functional limitations, such as advanced knee osteoarthritis or conditions that primarily impact lower-extremity

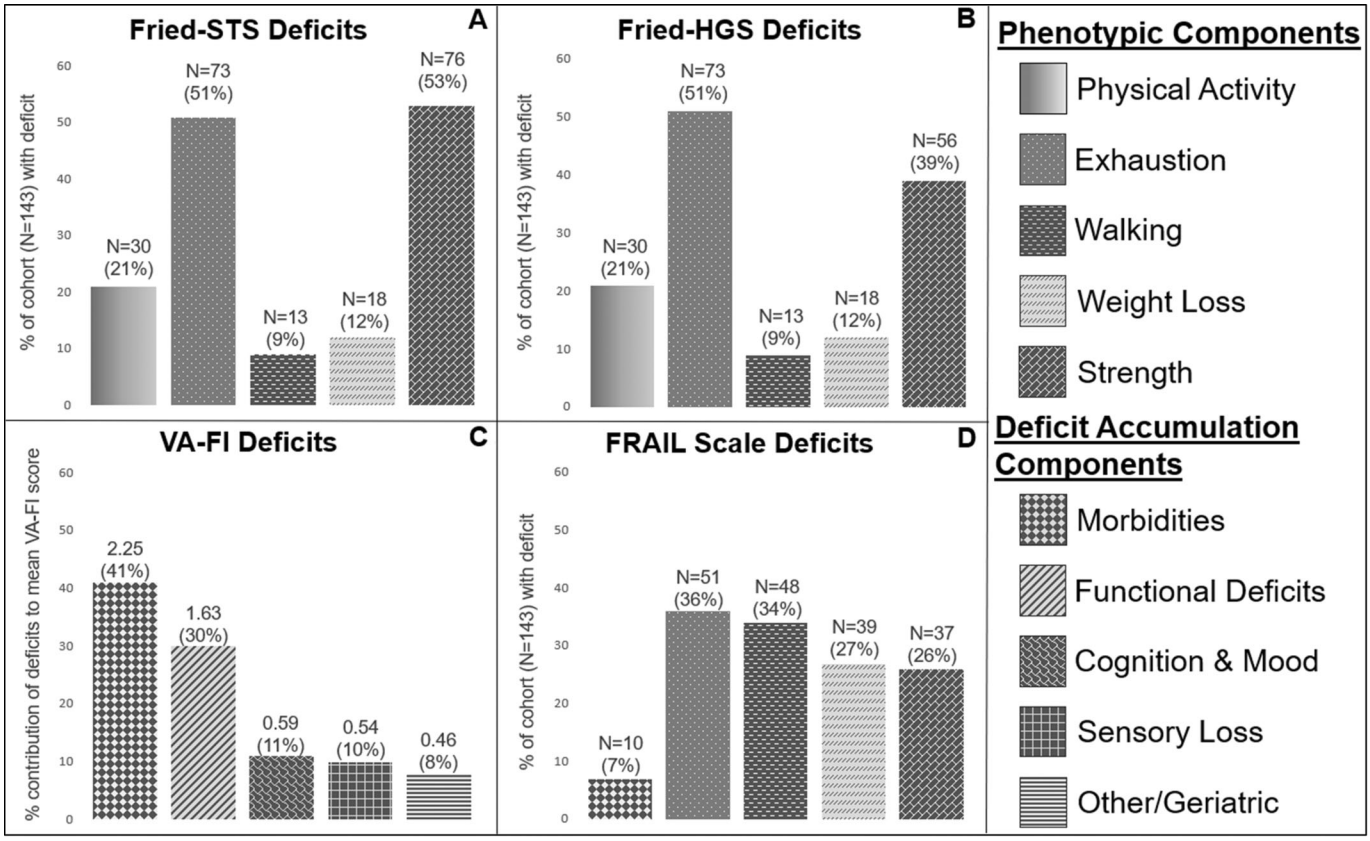

Figure 3. Prevalence of deficits included in each frailty measure for the entire cohort (N = 143). (A) Fried-STS deficits. (B) Fried-HGS deficits. (C) VA-FI deficits. (D) FRAIL Scale deficits. VA-FI components were calculated as the mean number of deficits (percentage contribution to mean score). The FRAIL Scale, Fried-STS, and Fried-HGS deficits were measured as n (%). The legend categorizes the components by frailty framework, with the pattern corresponding to a bar in the graph if that framework's component is included in the given frailty measure. Because many participants had more than one deficit, the sum of the deficit category percentages for each frailty measure is >100%. Fried-HGS, Fried Frailty Phenotype with handgrip strength; Fried-STS, Fried Frailty Phenotype with chair sit-to-stand; VA-FI, Veterans Affairs Frailty Index.

muscle function (eg, stroke or those who are wheelchair bound for other reasons). Future studies are needed to explore the relationship of the Fried-STS with poor health outcomes in a larger RA cohort with more diverse health outcomes available.

Prior studies comparing frailty instruments in RA showed moderate to high agreement[33,34]; however, these studies used frailty instruments with highly overlapping constructs. The predictive ability of multiple frailty instruments on health outcomes in RA has been examined in one previous study[35,36] of two separate RA cohorts. In this study, Hanlon et al used a frailty index and an adapted phenotypic instrument, both of which predicted death and hospitalization.[36] Adding to this prior work, we compared four unique instruments: two phenotypic instruments (Fried-STS and Fried-HGS), one deficit accumulation model (VA-FI), and one patient-reported mixed model (FRAIL Scale). Although the FRAIL Scale was designed to have construct overlap with phenotypic frailty measures as well as deficit accumulation models,[37] we found low agreement by kappa testing between instruments in our cohort. Despite poor overall agreement between instruments of different conceptual frameworks, we found that each instrument predicted poor health outcomes to varying degrees. Frailty by the VA-FI, driven by morbidities, was associated with hospitalizations, whereas frailty by Fried and FRAIL Scale measures, driven by exhaustion and physical function deficits, was associated with falls. Despite these differences, all measures, except the Fried-HGS, were associated with the composite outcome. Notably, the FRAIL Scale was able to predict outcomes similar to the objectively measured Fried instruments, further validating the utility of patient-reported outcomes. Although exploratory, our findings are the first to our knowledge to use three validated frailty instruments, as well as a novel pragmatic alternative, when evaluating adverse health outcomes in RA.

The primary limitation of this study is the size of the cohort, resulting in a limited number of health outcomes for our exploratory analysis (eg, only two deaths during follow-up). It is also possible that the Fried-STS predicted falls because STS reflects lower-extremity function. It is notable that the Fried-STS maintained its significant association with the composite outcome; however, falls contributed to the majority of the composite outcome. We recognize that self-reported falls will be subject to

Table 2. Baseline demographics, clinical characteristics, and long-term outcomes of participants with follow-up data (N = 136) and individuals captured as frail by each measure*

		Frail participants by each frailty instrument			
	Total (N = 136)	Fried-STS (n = 23, 17%)	Fried-HGS (n = 21, 15%)	VA-FI (n = 49, 36%)	FRAIL Scale (n = 27, 20%)
Sociodemographic					
Age, mean (SD), y	64.4 (11.9)	67.0 (8.7)	67.7 (8.7)	68.5 (9.8)	67.5 (11.7)
Male sex, n (%)	100 (74)	18 (78)	19 (91)	36 (74)	24 (89)
Race, n (%)					
Black	17 (13)	0	0	5 (10)	0
Multiple races	16 (12)	3 (13)	4 (19)	6 (12)	4 (15)
Other[a]	9 (6)	2 (9)	0 (0)	2 (4)	0 (0)
White	94 (69)	18 (78)	17 (81)	36 (74)	23 (85)
Ethnicity, n (%)					
Hispanic/Latino	8 (6)	0	0	0	2 (7)
BMI, mean (SD)	30.1 (5.5)	28.8 (5.7)	28.7 (6.0)	30.0 (5.4)	31.7 (5.4)
Smoking, mean (SD), pack years	20.1 (30.3)	36.4 (43.5)	38.2 (45.7)	24.8 (31.7)	24.7 (23.1)
RA disease characteristics					
Disease duration, mean (SD), y	14.6 (13.7)	11.8 (10.1)	13.5 (13.6)	14.4 (13.6)	15.9 (16.4)
MDHAQ score (range 0–3.0), mean (SD)	0.7 (0.5)	1.1 (0.4)	1.1 (0.4)	0.8 (0.4)	1.1 (0.4)
Pain score (range 0–10), mean (SD)	4.3 (2.5)	5.9 (2.1)	5.9 (2.6)	5.0 (2.4)	6.1 (2.1)
CCP positive, n (%)	90 (66)	14 (61)	12 (57)	33 (67)	18 (67)
DAS28-ESR (range 0–9.4), mean (SD)	3.7 (1.4)	4.4 (1.2)	4.5 (1.3)	4.1 (1.1)	4.7 (1.5)
Erosions, n (%)	55 (40)	10 (44)	9 (43)	21 (43)	11 (41)
Deformities, n (%)	44 (32)	8 (35)	9 (43)	19 (39)	7 (26)
Joint replacement, n (%)	30 (22)	5 (22)	6 (29)	13 (27)	9 (33)
Poor health outcomes					
Number of falls, mean (SD)	1.4 (3.1)	2.9 (4.1)	2.6 (4.3)	2.5 (4.7)	2.9 (4.2)
0 falls, n (%)	81 (60)	9 (39)	12 (57)	25 (51)	13 (48)
1 fall, n (%)	15 (11)	1 (4)	0 (0)	4 (8)	1 (4)
>1 fall, n (%)	40 (29)	13 (57)	9 (43)	20 (41)	13 (48)
Cumulative days hospitalized, mean (SD)	1.5 (5.4)	4.5 (11.4)	4.1 (11.5)	3.8 (8.5)	3.2 (9.1)
0 days, n (%)	113 (83)	16 (70)	15 (71)	33 (67)	20 (74)
1–3 days, n (%)	6 (4)	3 (13)	2 (10)	3 (6)	2 (7)
>3 days, n (%)	17 (13)	4 (17)	4 (19)	13 (27)	5 (19)
Composite incidence of fall, hospitalization, or death, n (%)	65 (48)	15 (65)	10 (48)	31 (63)	18 (67)
Fall (yes/no)	55 (40)	14 (61)	9 (43)	24 (49)	14 (52)
Hospitalization (yes/no)	22 (16)	6 (26)	5 (24)	15 (31)	7 (26)
Died	2 (2)	2 (9)	1 (5)	1 (2)	1 (4)

* BMI, body mass index; CCP, cyclic citrullinated peptide; DAS28-ESR, Disease Activity Score in 28 joints using the erythrocyte sedimentation rate, HGS, handgrip strength; MDHAQ, Multidimensional Health Assessment Questionnaire; RA, rheumatoid arthritis; STS, sit-to-stand; VA-FI, Veterans Affairs Frailty Index.
[a] Other race: American Indian, Alaskan Native, Pacific Islander, Asian, other, and unknown.

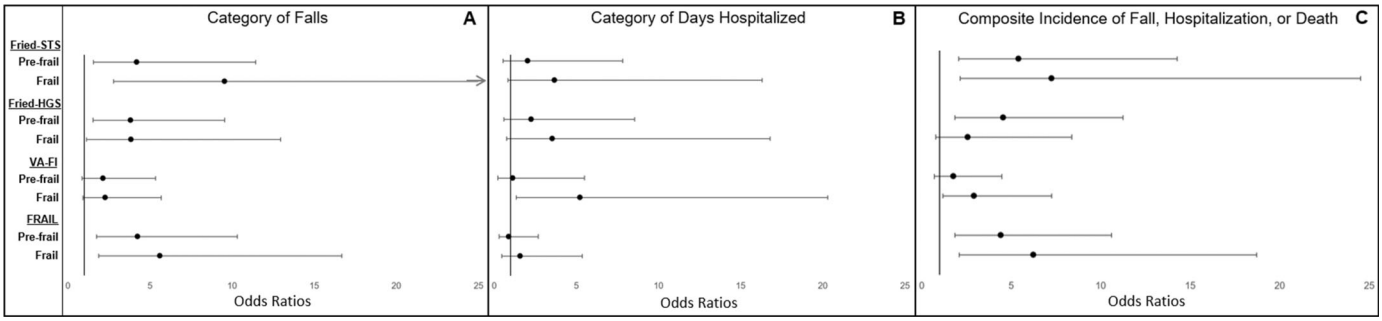

Figure 4. Forest plots for results of adjusted logistic and ordinal logistic regression models for each of the four frailty measures. Models were adjusted for age and sex. Robust frailty category was the referent for all models. (A) Ordinal logistic regression of baseline frailty and category of falls (none, one, more than one) over one year of follow-up. (B) Ordinal logistic regression of baseline frailty and category of cumulative days of hospitalization (none, one to three days, more than three days) over one year of follow-up. (C) Logistic regression of baseline frailty and incidence of an adverse outcome (fall, hospitalization, or death). Fried-HGS, Fried Frailty Phenotype with handgrip strength; Fried-STS, Fried Frailty Phenotype with chair sit-to-stand; VA-FI, Veterans Affairs Frailty Index.

some level of recall bias. However, we collected data on falls from participants on a six-month interval basis to minimize that bias. Our study required in-person visits and a life expectancy greater than two years to allow for follow-up, which may have limited our ability to recruit the frailest participants. Lastly, our cohort is from a single center with a high proportion of men with established RA disease, which limits generalizability to all individuals with RA. Due to our single-center design, limited follow-up data, and small sample size, we acknowledge the importance of replicating these results to further investigate the use of the Fried-STS in larger external RA cohorts. This study had many strengths. The same two rheumatologists and study staff conducted all research examinations, amplifying precision and limiting variability. The male predominance adds important knowledge as male patient are underrepresented in RA research. Additionally, the veteran-only population allows us to address unique health concerns of veterans and provides us with comprehensive data from VA electronic health records and administrative data. Our cohort, although smaller, is focused on frailty, so we have comprehensive data to directly compare three validated, unique frailty instruments against a novel pragmatic alternative that uses a strength that has been validated for frailty measurement in other disease states but has not yet been used in RA.[3,9]

As expected, the four frailty instruments (Fried-STS, Fried-HGS, VA-FI, and FRAIL Scale) demonstrated low agreement in classifying frailty in our cohort. Each instrument was able to predict poor health outcomes, with phenotypic and patient-reported instruments predicting falls and the composite outcome and the deficit accumulation model predicting hospitalizations and the composite measure. The Fried-HGS demonstrated the lowest precision in its ability to predict poor health outcomes among the frailty instruments evaluated. Additionally, our data suggest that special consideration should be given to the use of HGS as a measurement of whole-body strength in RA, particularly in those with long-standing disease and deformities. Overall, our findings support the idea that the best frailty measure is the one available, but the individual's health conditions and the outcome of interest should be considered when deciding on which instrument to use.[38] Further investigation of frailty instruments in larger RA cohorts is needed, particularly to evaluate a wider array of health outcomes while also accounting for the impacts of RA disease.

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 Wysham 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. Fried LP, Tangen CM, Walston J, et al; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci 2001;56(3):M146–M157. doi:10.1093/gerona/56.3.M146
2. Kim DH, Rockwood K. Frailty in older adults. N Engl J Med 2024; 391(6):538–548. doi:10.1056/NEJMra2301292
3. Loecker CN, Schmaderer MS, Wysham KD, et al. Psychometric properties of frailty instruments in adults with rheumatoid arthritis: a systematic review. ACR Open Rheumatol 2024;6(2):91–102. doi:10.1002/acr2.11640
4. Haider S, Grabovac I, Berner C, et al. Frailty in seropositive rheumatoid arthritis patients of working age: a cross-sectional study. Clin Exp Rheumatol 2019;37(4):585–592.
5. Germonpré S, Mulier S, Falzon L, et al. Prevalence of frailty and pre-frailty in patients with rheumatoid arthritis: a systematic literature review and meta-analysis. Clin Exp Rheumatol 2023;41(7):1443–1450. doi:10.55563/clinexprheumatol/1kr7bs
6. Gao RC, Wu ZG, Wu ZZ, et al. Frailty in rheumatoid arthritis: a systematic review and meta-analysis. Joint Bone Spine 2022;89(4):105343. doi:10.1016/j.jbspin.2022.105343
7. England BR, Roul P, Yang Y, et al. Burden and trajectory of multimorbidity in rheumatoid arthritis: a matched cohort study from 2006 to 2015. Ann Rheum Dis 2021;80(3):286–292. doi:10.1136/annrheumdis-2020-218282
8. Cook MJ, Verstappen SMM, Lunt M, et al. Increased frailty in individuals with osteoarthritis and rheumatoid arthritis and the influence of comorbidity: an analysis of the UK Biobank cohort. Arthritis Care Res (Hoboken) 2022;74(12):1989–1996. doi:10.1002/acr.24747
9. Hanlon P, Morrison H, Morton F, et al. Frailty in people with rheumatoid arthritis: a systematic review of observational studies. Wellcome Open Res 2022;6:244. doi:10.12688/wellcomeopenres.17208.2
10. Lacas A, Rockwood K. Frailty in primary care: a review of its conceptualization and implications for practice. BMC Med 2012;10(1):4. doi:10.1186/1741-7015-10-4
11. Theou O, Brothers TD, Mitnitski A, et al. Operationalization of frailty using eight commonly used scales and comparison of their ability to predict all-cause mortality. J Am Geriatr Soc 2013;61(9):1537–1551. doi:10.1111/jgs.12420
12. Fried LP, Cohen AA, Xue QL, et al. The physical frailty syndrome as a transition from homeostatic symphony to cacophony. Nat Aging 2021;1(1):36–46. doi:10.1038/s43587-020-00017-z
13. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of deficits as a proxy measure of aging. ScientificWorldJournal 2001;1:323–336. doi:10.1100/tsw.2001.58
14. Wolfe F, Hawley DJ, Wilson K. The prevalence and meaning of fatigue in rheumatic disease. J Rheumatol 1996;23(8):1407–1417.
15. 15. Dedeoğlu M, Gafuroğlu Ü, Yilmaz Ö, et al. The relationship between hand grip and pinch strengths and disease activity, articular damage, pain, and disability in patients with rheumatoid arthritis. Arch Rheumatol 2013;28(2):69–77. doi:10.5606/tjr.2013.2742
16. Bohannon RW, Bubela DJ, Magasi SR, et al. Sit-to-stand test: performance and determinants across the age-span. Isokinet Exerc Sci 2010;18(4):235–240. doi:10.3233/IES-2010-0389
17. Kim M, Won CW. Cut points of chair stand test for poor physical function and its association with adverse health outcomes in community-dwelling older adults: a cross-sectional and longitudinal study. J Am Med Dir Assoc 2022;23(8):1375–1382.e3. doi:10.1016/j.jamda.2021.11.007
18. Manay P, Ten Eyck P, Kalil R, et al. Frailty measures can be used to predict the outcome of kidney transplant evaluation. Surgery 2021; 169(3):686–693. doi:10.1016/j.surg.2020.07.016

19. Lai JC, Covinsky KE, McCulloch CE, et al. The liver frailty index improves mortality prediction of the subjective clinician assessment in patients with cirrhosis. Am J Gastroenterol 2018;113(2):235–242. doi:10.1038/ajg.2017.443
20. Orkaby AR, Nussbaum L, Ho YL, et al. The burden of frailty among U.S. veterans and its association with mortality, 2002-2012. J Gerontol A Biol Sci Med Sci 2019;74(8):1257–1264. doi:10.1093/gerona/gly232
21. Abellan van Kan G, Rolland Y, Bergman H, et al. The I.A.N.A Task Force on frailty assessment of older people in clinical practice. J Nutr Health Aging 2008;12(1):29–37. doi:10.1007/BF02982161
22. Aletaha D, Neogi T, Silman AJ, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010;62(9):2569–2581. doi:10.1002/art.27584
23. Rosenthal MA, Ranji SR, Kanzaria HK, et al. Characterizing patients hospitalized without an acute care indication: a retrospective cohort study. J Hosp Med 2023;18(4):294–301. doi:10.1002/jhm.13061
24. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. J Youth Adolesc 1991;20(2):149–166. doi:10.1007/BF01537606
25. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381–1395. doi:10.1249/01.MSS.0000078924.61453.FB
26. Cruz-Jentoft AJ, Bahat G, Bauer J, et al; Writing Group for the European Working Group on Sarcopenia in Older People 2 (EWGSOP2), and the Extended Group for EWGSOP2. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019;48(4):601. doi:10.1093/ageing/afz046
27. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum 1999;42(10):2220–2230. doi:10.1002/1529-0131(199910)42:10<2220:AID-ANR26>3.0.CO;2-5
28. Prevoo ML, van 't Hof MA, Kuper HH, et al. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38(1):44–48. doi:10.1002/art.1780380107
29. Stata Statistical Software: Release 18. StataCorp LLC; 2023.
30. Arab Alkabeya H, Hughes AM, Adams J. Factors associated with hand and upper arm functional disability in people with rheumatoid arthritis: a systematic review. Arthritis Care Res (Hoboken) 2019;71(11):1473–1481. doi:10.1002/acr.23784
31. Brubeck HF, Riggles KE, Bass RS, et al. Evaluating the longitudinal association of rheumatoid arthritis disease activity with phenotypic frailty: evidence for secondary frailty? Arthritis Care Res (Hoboken) 2026;78(3):307–315. doi:10.1002/acr.25615
32. Sobue Y, Suzuki M, Ohashi Y, et al. Validation of grip strength as a measure of frailty in rheumatoid arthritis. Sci Rep 2022;12(1):21090. doi:10.1038/s41598-022-21533-5
33. Oetsma S, Boonen A, Starmans M, et al. Validation of two frailty questionnaires in older patients with rheumatoid arthritis: a cross-sectional study. Clin Exp Rheumatol 2020;38(3):523–528.
34. Salaffi F, Farah S, Di Carlo M. Frailty syndrome in rheumatoid arthritis and symptomatic osteoarthritis: an emerging concept in rheumatology. Acta Biomed 2020;91(2):274–296. doi:10.23750/abm.v91i2.9094
35. Li G, Chen M, Li X, et al. Frailty and risk of osteoporotic fractures in patients with rheumatoid arthritis: data from the Ontario Best Practices Research Initiative. Bone 2019;127:129–134. doi:10.1016/j.bone.2019.06.006
36. Hanlon P, Morton F, Siebert S, et al. Frailty in rheumatoid arthritis and its relationship with disease activity, hospitalisation and mortality: a longitudinal analysis of the Scottish Early Rheumatoid Arthritis cohort and UK Biobank. RMD Open 2022;8(1):e002111. doi:10.1136/rmdopen-2021-002111
37. Abellan van Kan G, Rolland YM, Morley JE, et al. Frailty: toward a clinical definition. J Am Med Dir Assoc 2008;9(2):71–72. doi:10.1016/j.jamda.2007.11.005
38. Singh N, Faye AS, Abidi MZ, et al. Frailty integration in medical specialties: current evidence and suggested strategies from the Clin-STAR frailty interest group. J Am Geriatr Soc 2025;73(4):1029–1040. doi:10.1111/jgs.19268

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1177–1187
DOI 10.1002/acr.80000
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

American College of Rheumatology
Empowering Rheumatology Professionals

Addressing Economic Insecurities Can Improve Patient-Reported Outcomes in Lupus

Jay Patel, BS,[1] Tripti Singh, MD,[1] Meredith Ingersoll, MSW,[1] Shelby Gomez, PharmD,[1] Amanda Weber, RN, BSN,[1] Sarah E. Panzer, MD,[1] Sancia Ferguson, MD,[1] Christie M. Bartels, MD, MS,[1] and Shivani Garg, MD, PhD[1,2]

Objective. Economic insecurities, such as food, housing, transportation, and financial challenges, are modifiable risk factors and influence patient-reported outcomes (PROs) in systemic lupus erythematosus (SLE). We examined the following: (1) associations between economic insecurities and PROs, and (2) the impact of screening and addressing economic insecurities during SLE visits.

Methods. In the Collaborative Lupus Clinics at the University of Wisconsin–Madison, patients were routinely screened for economic insecurities and met with a social worker (SW) during visits. Clinical data including PROs from the Patient-Reported Outcomes Measurement Information System Global Health Short Form at baseline and follow-up were abstracted from the Collaborative Lupus Clinics Data Repository. Using multivariable linear regression, associations among economic insecurities, social drivers of health (eg, insurance), and PROs were assessed. Next, changes in PROs following SW discussions were evaluated.

Results. Among 222 patients (mean age 47 years; 90% women), 16% reported at least one economic insecurity. Each 1-point increase in economic insecurity score was linked with lower PRO T scores in all domains: physical health (–1.85, $P = 0.04$), mental health (–1.32, $P = 0.17$), and social function (–0.24, $P = 0.03$). A sequential increase in economic insecurities, at least one, at least two, and at least three, lowered physical health by 2.00, 3.86, and 9.10 points, respectively. Patients with economic insecurities and Medicaid/no insurance had two times lower PRO scores in all domains. Following SW intervention, PROs improved by 4.52, 1.12, and 0.69 points in all domains.

Conclusion. Although economic insecurities negatively affect PROs in SLE, a systematic approach to assess and address economic insecurities in clinics can improve PROs over time in SLE.

INTRODUCTION

Systemic lupus erythematosus (SLE; or lupus) is the most common life-threatening autoimmune disease.[1] Individuals with lupus face a 20-fold higher risk of end-stage kidney disease, a six-fold higher risk of cardiovascular disease, and a six-fold higher risk of premature death.[2–5] Likewise, individuals with lupus often have lower than average scores in patient-reported outcomes (PROs), a reflection of an individual's self-evaluation of health due to chronic disease. Poor PROs in lupus could likely explain 31% more work disability compared to other diseases.[6] However, biologic factors and genetic variability cannot fully account for the heterogeneity in clinic outcomes and PROs.[7] To improve lupus outcomes, it is important to understand how social drivers of health (SDoH) impact clinical outcomes and PROs in lupus.

Our group's meta-analysis reported a multiplicative negative interaction two or more adverse SDoH driving 12-fold higher odds of poor clinical outcomes in lupus nephritis (LN).[8] However, similar associations among SDoH, biologic factors, and PROs have not been established across geographically different population groups in lupus. A recent study using the California Lupus (CLUES) cohort highlighted that economic insecurities (ie, food, housing, transportation, financial insecurity) and not conventionally measured SDoH (ie, employment, insurance) were associated with poor PROs.[9] However, there are stark differences in SLE presentation, SDoH, economic insecurities, insurance coverage, and health care access in the Midwest region compared to the

This work was presented at the 2024 American College of Rheumatology Convergence Meeting as an oral presentation.
[1]School of Medicine and Public Health, University of Wisconsin–Madison, Madison, Wisconsin. [2]Yale School of Medicine, New Haven, Connecticut.
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.80000).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80000.
Address correspondence via email to Shivani Garg, MD, PhD, at shivani.garg@yale.edu.
Submitted for publication July 4, 2025; accepted in revised form December 17, 2025.

SIGNIFICANCE & INNOVATIONS

- This study validates economic insecurities as a distinct risk factor associated with worse patient-reported outcomes (PROs) in systemic lupus erythematosus (SLE).
- A sequential negative impact on all PRO domains is seen with increasing economic insecurities in SLE, and having economic insecurities combined with other adverse social drivers of health (SDoH) is associated with a two-fold negative impact on PROs.
- This study highlights a systematic approach to using a validated electronic health record–based screening tool (Epic SDoH Snapshot) in clinics for assessing economic insecurities.
- This is the first study to propose a scalable model for integrating social work intervention in SLE clinics and a positive impact of assessing and addressing economic insecurities on improving PROs in all domains over six months' follow-up in SLE.

West.[10,11] Therefore, there is a need to validate associations between PROs and SDoH in other cohorts to enhance generalizability. Additionally, data on how SDoH should be assessed and addressed during a clinic visit are lacking. Finally, it is important to estimate the impact of addressing modifiable SDoH on PROs and clinical outcomes in lupus. Such data are critical for identifying key factors among SDoH domains, the interactions among such factors, potentially modifiable targets, and guidance on how to intervene to improve outcomes, particularly PROs, in lupus.

Therefore, leveraging data from the Collaborative Lupus Clinics' Data Repository at the University of Wisconsin–Madison (UW),[12] we aimed to examine associations between economic insecurities and conventional SDoH with PROs in lupus. Additionally, we tested interactions between economic insecurities and conventional SDoH and the impact of such interactions on PROs to identify high-risk groups and high-risk social factors. Finally, we assessed the longitudinal impact of a social worker (SW)-based intervention, which has been integrated into our Collaborative Lupus Clinics as part of routine care,[12] on addressing economic insecurities and improving PROs in our cohort.

METHODS

Collaborative Lupus Clinics teams at the UW. We have established multidisciplinary collaborative lupus and LN clinics at the UW. The UW Collaborative Lupus Clinics care for more than 400 unique patients and provide over 700 patient visits annually, serving statewide. Approximately 35% of patients identify as Asian or Black/African American. These teams comprise rheumatologists, nephrologists, an SW, skilled medical staff, and pharmacists. We have created workflows to screen and address SDoH as part of routine lupus care in our clinics. More details about the multidisciplinary clinics can be found in our 2022 manuscript.[12]

Routine clinical workflow for SDoH screening and SW-driven intervention in our Collaborative Lupus Clinics. We developed a workflow to screen for social risk factors (including economic insecurities) as part of routine lupus care, using the Epic electronic health record (EHR). The workflow includes four steps (Figure 1A).

Step 1: previsit screening. One week before the visit, as part of our previsit screening, a medical assistant on the lupus clinic team calls all new and established patients with a diagnosis of lupus or LN who are scheduled to be seen in the UW's Collaborative Lupus Clinics. During the call, the medical assistant screens for social barriers that are currently present and confirms that the laboratory tests are completed before the scheduled visit. We used the existing SDoH screening tool, available within the Epic EHR and in concordance with the guidelines from the National Academy of Medicine (NAM) and Center for Medicare and Medicaid Services (CMS), to screen for economic insecurities in our clinics.[13,14] To elaborate, the SDoH screening tool that we used uses validated screening questions to identify four different economic insecurities including transportation barriers, financial barriers, food insecurities, and unstable housing. Details on validated questions used to screen for these economic insecurities are listed in Figure 1B and Supplementary Table 1. Positive screens for any social barriers were shared with our SW in UW Collaborative Lupus Clinics. The SW contacted patients with at least one social barrier to address any barriers that could otherwise result in no-shows or cancellations (Figure 1A).

Step 2: during clinic visit. For patients with any economic insecurities, the SW met with them during their lupus clinic visit and developed a personalized plan with the patients to address their barriers. This plan included sharing resources such as providing food stamps, setting up rides for medical visits, applying for financial assistance, or connecting with the pharmacist for copay or medicine coverage issues (Figure 2).

Step 3: after clinic visit follow-up. The SW connected with the patient via phone or a message within 8 to 12 weeks after the visit to ensure that the barriers were addressed or resolved. If economic insecurities were resolved, then the EHR questionnaire was updated.

Step 4: rescreening for SDoH. For patients who screened positive for any social barriers or economic insecurities in the last three to six months, they were rescreened one week before their three- to six-month follow-up visit as part of previsit screening (step 1). Patients who had never reported any social barriers were only screened once a year in our clinics.

Study design. This is a medical records review observational study that used data from the Collaborative Lupus Clinics'

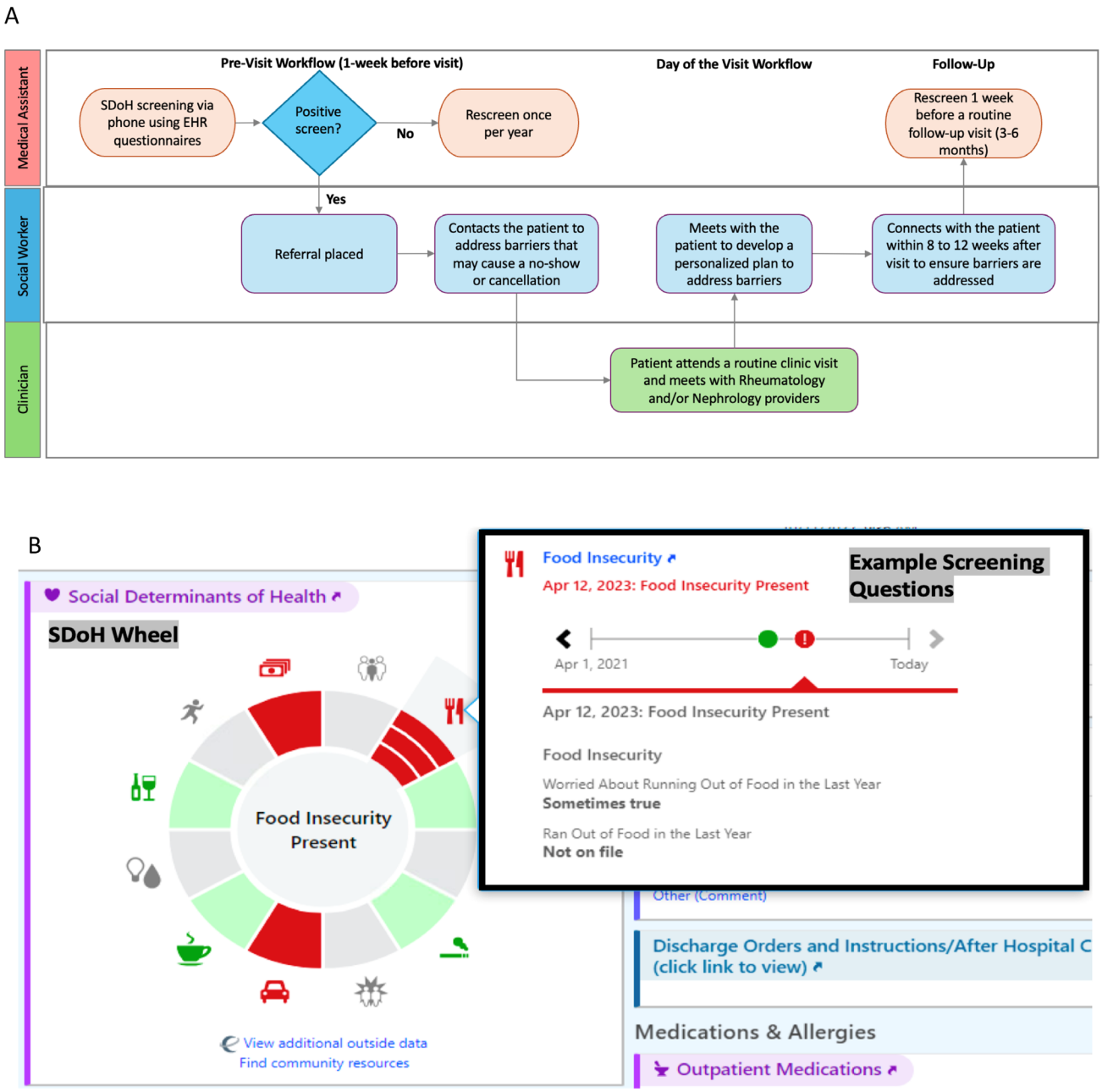

Figure 1. Clinical workflow to assess and address economic insecurities at baseline lupus visit. (A) Social work intervention workflow in our Collaborative Lupus Clinics at the University of Wisconsin–Madison. (B) The screening tool with the Epic SDoH Snapshot, part of Epic's Healthy Planet module, supports social risk factor screening, including food insecurity, transportation needs, and financial insecurity. EHR, electronic health record; SDoH, social drivers of health.

Data Repository at the UW in 2020 to 2024. The participants in this Collaborative Lupus Clinics' Data Repository met the following eligibility criteria: (1) residency in Wisconsin, Illinois, or neighboring states; (2) age ≥18 years; (3) the ability to provide informed consent for the study team to access medical records for data abstraction over time; and (4) a validated diagnosis of SLE according to the 1997 American College of Rheumatology (ACR) SLE classification criteria,[15] 2012 Systemic Lupus International Clinics (SLICC) classification criteria,[16] or 2019 EULAR/ACR classification for SLE.[17] The study protocol was approved by the UW School of Medicine and Public Health Institutional Review Board (IRB; 2019-0942).

Data collection. At every lupus clinic visit, the following data were completed as part of clinical care: the Patient-Reported Outcomes Measurement Information System (PROMIS) Global Health Short Form completed by patients as a PRO measure,[18] patient-reported medication adherence, and clinician-reported

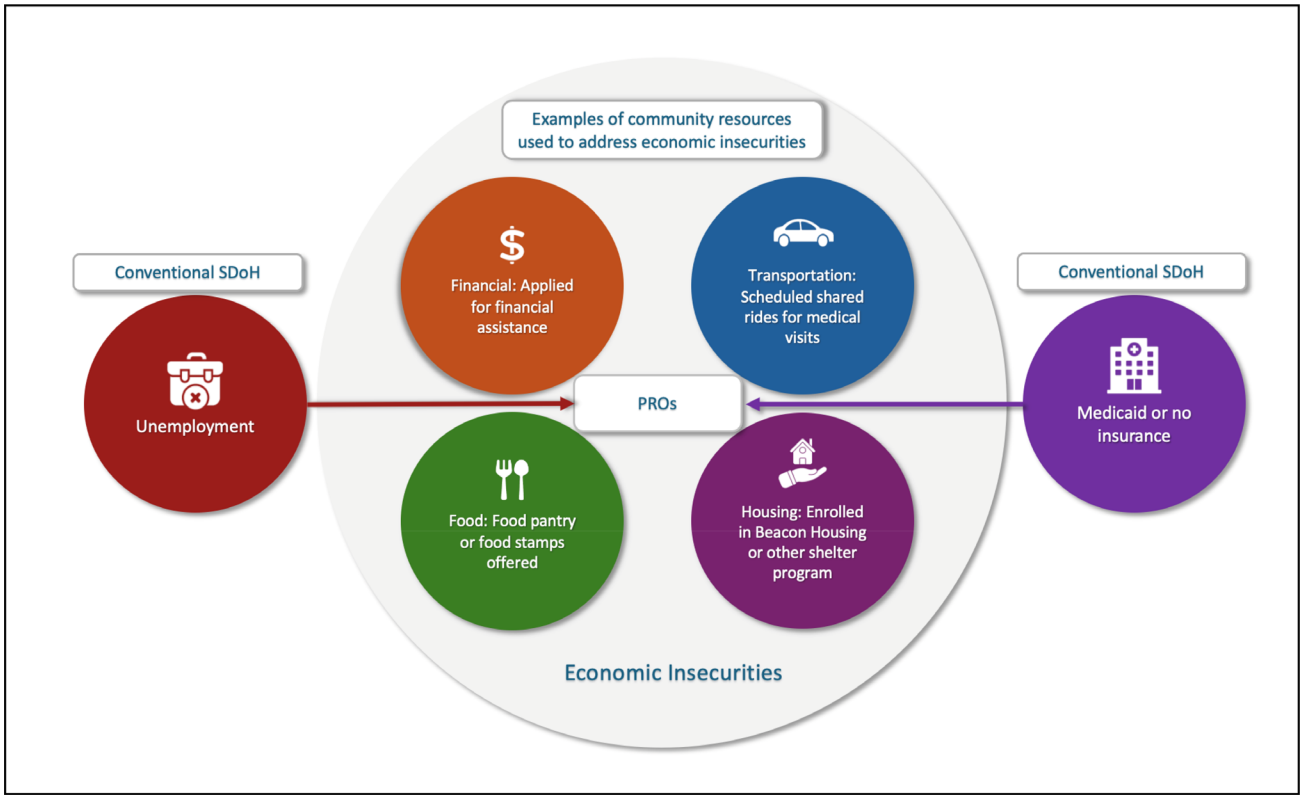

Figure 2. A model showing modifiable social risk factors (economic insecurities) and other social drivers of poor patient-reported outcomes (PROs) in lupus and how economic insecurities can be addressed, leveraging community resources and social workers' support in clinics. SDoH, social drivers of health. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80000/abstract.

SLE Disease Activity Index 2000 (SLEDAI-2K)[19] and SLICC Damage Index (SDI).[20] Data were manually abstracted by a trained research coordinator for a priori defined clinical variables (IRB 2019-0942) from each consented patient's EHR every quarter. Periodic reviews were conducted by the study team to ensure data reliability and completeness.

Variables. *Key explanatory variable: economic insecurities.* From the index visit date, we abstracted data on economic insecurities, including housing instability, food insecurity, financial insecurity (eg, difficulty affording utilities, medical care, housing), and transportation barriers (eg, missing medical appointments or difficulty obtaining medications). Given a person could experience multiple economic insecurities, we used economic insecurities as a continuous variable in which each additional economic insecurity was represented as a 1-unit increase.[9] In a separate model, economic insecurities were categorized as at least one, at least two, and at least three economic insecurities being present.

Other variables: conventional SDoH. From the index visit date, we also abstracted data on conventional SDoH domains including individual, health behaviors, and community. These data included employment status, which was categorized as unemployed, employed, student, or retired. Next, insurance was categorized as Medicaid/no insurance versus commercial or private insurance. Medicaid eligibility or uninsured status grouping served as a surrogate for low personal income. For community SDoH domain, we calculated Area Deprivation Index (ADI) and categorized it as <80% decile versus ≥80% decile, indicating neighborhoods at highest risk of disadvantage. We also abstracted data on high-risk behaviors, (smoking, recreational drug use, alcohol use, etc). Finally, given mental health factors such as patient-reported loneliness or depression could affect PROs, this information was also abstracted.[8]

Covariables: sociodemographic and other factors. From the index clinic visit, we abstracted key demographics and SLE-specific variables from the participants' medical records. Demographics included age (per 10-year increase), patient-reported sex (male vs female), and race or ethnicity (White, Black or African American, Hispanic or Latino, Asian). SLE-specific variables included SLEDAI-2K as a continuous variable (per 1-unit increase),[19] SDI as a continuous variable (1-unit increase),[20] and duration of SLE in years. Next, we abstracted data on immunosuppressive medications, prednisone dose at the visit in milligrams (per 1-mg increase), and hydroxychloroquine (HCQ) whole blood (WB) levels (>750 vs ≤750 ng/mL) as a proxy for effective HCQ use.[21–23]

Primary outcome: PROs. From the index clinic visit and three- to six-month follow-up clinic visit, PROs of physical health, mental health, and social function were collected from the PROMIS Global Health Short Form (10 questions).[18] Raw scores for PROs for physical health, mental health, and social function were calculated and converted to normalized T scores using the standard validated PROMIS scoring guidelines.[24] Normalized T scores for PROs for physical health, mental health, and social function were used as continuous variables.

Statistical methods. *Cross-sectional analysis using data from index clinic visit.* We first performed descriptive analysis and summarized the baseline characteristics of our cohort using means and standard deviations for continuous variables and frequencies for categorical variables.

Next, unadjusted linear regression analysis was completed to understand the association between each variable, including demographics, SLE-specific variables, medications, conventional SDoH, and economic insecurities, with PRO T scores for physical health, mental health, and social function. PRO T scores for physical health, mental health, and social function were analyzed in separate models to understand how different factors affect PROs in each domain.

Then, multivariate linear regression analyses were performed to examine associations among the key explanatory variable, economic insecurities, and PRO T scores in physical health, mental health, and social function adjusting for covariates. Again, separate models were built to analyze associations or drivers of PRO T scores in each domain. Our multivariable models were adjusted for age, sex, race or ethnic groups, SLEDAI score, SDI score, prednisone dose, ADI score, HCQ WB levels, high-risk health behavior, and patient-reported loneliness or depression. Each PRO (physical health, mental health, and social function) was modeled as a separate continuous outcome variable, and results were reported as mean difference or change in mean.

Additionally, to understand the sequential impact of increasing economic insecurities on PROs, we used economic insecurities as a categorical variable and the following categories were used in the analysis: at least one economic insecurity, at least two economic insecurities, and at least three economic insecurities. For each category of economic insecurity, we performed multivariable linear regression analysis to calculate the mean change in PRO in each domain (physical health, mental health, social function) with versus without economic insecurities. To assess the linear trend across increasing levels of economic insecurity on each domain, we performed a *P* for trend analysis.

Finally, we performed three subgroup analyses to test interactions between economic insecurities and conventional SDoH (insurance status, Medicaid/no insurance vs commercial insurance; employment status, unemployed; race, Black or African American). These groups were selected due to their established association with worse clinical outcomes.[8] Given limited sample size, reduced multivariable regression models were developed to assess how economic insecurities resulted in mean change in PRO in each subgroup, adjusting for only key covariables.

Longitudinal analysis comparing change in PROs at three- to six-month follow-up visit to baseline visit. Using a pre–post multivariable linear regression model, we assessed if addressing or resolving economic insecurities with the help of community resources or SW resulted in a change in PROs in each domain. We included both unadjusted and adjusted models. Variables adjusted for were age, race, sex, insurance status, ADI score, SLEDAI score at the follow-up visit, SDI score at the follow-up visit, prednisone dose, HCQ WB levels, and medications at the follow-up visit. Economic insecurities addressed (yes or no) was the key explanatory variable in this model.

Missing data handling. We imputed missing variables using median values derived from participants' clinic visits within six

Table 1. Baseline characteristics (n = 222 patients)*

Variables at clinic visit	Mean ± SD or n (%)
Age at enrollment, y	46.5 ± 16.9
Female	199 (90)
Race and ethnicity (patient reported)	
White	163 (73)
Black or African American	28 (13)
Asian	19 (9)
Hispanic or Latino	12 (5)
Insurance type	
Commercial or Medicare	191 (86)
No insurance or Medicaid	31 (14)
Employment status	
Employed	136 (61)
Unemployed	46 (21)
Retired	32 (14)
Student	8 (4)
Area Deprivation Index < 80th percentile	17 (8)
Body weight, kg	80 ± 21
SLE disease duration, y	5.6 ± 4.9
Glomerular filtration rate, mL/min/1.73 m^2	88 ± 27
Patient-reported loneliness or depression	14 (6)
High-risk health behavior	9 (4)
Economic insecurities	
None	186 (84)
At least one	36 (16)
At least two	11 (5)
Three or more	4 (2)
SLEDAI-2K score	2.8 ± 3.3
SDI score	0.89 ± 1.1
HCQ dose, mg	323 ± 83
Weight-based HCQ dose ≤5 mg/kg/d	186 (84)
Therapeutic HCQ whole blood levels >750 ng/mL	123 (55)
Other SLE medicines	
Disease-modifying agents	69 (31)
Biologics	15 (7)
Combination therapy	13 (6)
Prednisone use	69 (31)
Prednisone dose, mg	3.2 ± 9.0

* HCQ, hydroxychloroquine; SDI, Systemic Lupus International Collaborating Clinics Damage Index; SLEDAI-2K, Systemic Lupus Erythematosus Disease Activity Index 2000.

months before the index visit. We performed a sensitivity analysis in which we first excluded participants with missing outcome variables (PRO scores) and then used imputed data. Given similar results and only missing PROs from five visits, we used the imputed data for the final models to avoid sample size reduction.

Given an observational cohort, we did not calculate a sample size. Using power.t.test in R software, we estimated that a sample size of 222 would detect even a small change of 0.22 in mean PRO T scores given the power was set at 90% and a two-sided alpha set at 0.05. All analyses were conducted using R software (version 2023.12.1). Statistical significance was defined as $P < 0.05$, and all results were reported with 95% confidence intervals (CIs).

RESULTS

This study included 222 participants (mean ± SD age 46.5 ± 16.9 years), of whom 90% (n = 199) were female. Racial and ethnic distribution was 73% White (n = 163), 13% Black or African American (n = 28), 9% Asian (n = 19), and 5% Hispanic or Latino (n = 12). At baseline, conventional SDoH included 31 patients with Medicaid or no insurance (14%), and 46 patients were unemployed (21%). A total of 36 patients reported at least one economic insecurity including transportation barriers, housing instability, food insecurity, or financial insecurity (16%) during their baseline visit (Table 1). The mean SLEDAI score was 2.8 ± 3.3; SDI score was 0.89 ± 1.1. Mean T scores were as follows: physical health 43.8 ± 7.45, mental health 45.5 ± 7.87, and social function 3.10 ± 0.89. Patients with at least economic insecurity had lower T scores across all domains: physical health 40.7 ± 6.88, mental health 42.3 ± 7.94, and social function 2.72 ± 0.81. Other patient characteristics at baseline are shown in Table 1.

Associations between economic insecurities and PROs using baseline visit data. Unadjusted bivariate analyses showed that each additional economic insecurity was associated with lower PRO T scores in physical health (–2.72, 95% CI –4.23 to –1.22, $P < 0.001$), mental health (–2.62, 95% CI –4.22 to –1.03, $P = 0.001$), and social function (–0.34, 95% CI –0.53 to –0.17, $P = 0.0002$) domains, respectively (Supplementary Table 2). After adjusting for covariables, each additional economic insecurity was independently associated with a 1.85-point lower

Table 2. Multivariable linear regression showing adjusted associations among economic insecurities, conventional SDoH, and PROs (n = 222)*

	Physical health		Mental health		Social function	
Variables at visit	Adjusted mean change (95% CI)	*P* value	Adjusted mean change (95% CI)	*P* value	Adjusted mean change (95% CI)	*P* value
Age per 10-year increase	0.018 (−0.06 to 0.09)	0.64	0.05 (−0.02 to 0.14)	0.15	−0.0003 (−0.01 to 0.01)	0.94
Female	0.16 (−3.27 to 3.58)	0.92	−1.31 (−5.03 to 2.41)	0.49	−0.20 (−0.62 to 0.21)	0.34
Black or African American	**−3.33 (−0.24 to −6.43)**	**0.03**	**−3.65 (−0.29 to −7.01)**	**0.03**	−0.18 (−0.55 to 0.20)	0.36
Asian	1.37 (−2.29 to 5.03)	0.46	0.45 (−3.52 to 4.43)	0.82	0.04 (−0.41 to 0.48)	0.87
Hispanic or Latino	1.23 (−3.25 to 5.71)	0.59	1.28 (−3.58 to 6.14)	0.60	0.21 (−0.33 to 0.76)	0.44
Body weight per 1-kg increase	−0.04 (−0.09 to 0.01)	0.09	−0.02 (−0.07 to 0.03)	0.65	−0.002 (−0.01 to 0.004)	0.58
Medicaid insurance/no insurance	−0.87 (−3.90 to 2.17)	0.57	−1.35 (−4.65 to 1.94)	0.42	−0.18 (−0.54 to 0.19)	0.35
ADI <80% decile	−1.35 (−5.06 to 2.37)	0.48	−0.92 (−4.96 to 3.11)	0.65	−0.11 (−0.56 to 0.34)	0.63
Loneliness or depression	2.36 (−2.00 to 6.72)	0.29	−1.65 (−6.39 to 3.08)	0.49	0.16 (−0.37 to 0.69)	0.55
High-risk health behavior per 1-unit increase	−0.56 (−5.69 to 4.57)	0.83	−2.30 (−7.87 to 3.27)	0.42	−0.13 (−0.75 to 0.49)	0.68
Employment						
Unemployed	**−3.09 (−0.31 to −5.87)**	**0.03**	−2.55 (−5.56 to 0.47)	0.10	**−0.49 (−0.16 to −0.83)**	**0.004**
Retired	−2.01 (−5.40 to 1.39)	0.25	0.42 (−3.27 to 4.10)	0.82	0.14 (−0.27 to 0.56)	0.49
Student	−0.99 (−6.90 to 4.92)	0.74	1.75 (−4.67 to 8.17)	0.59	−0.05 (−0.80 to 0.67)	0.89
Economic insecurities per 1-unit increase	**−1.85 (−0.11 to −3.58)**	**0.04**	−1.32 (−3.20 to 0.56)	0.17	**−0.24 (−0.03 to −0.45)**	**0.03**
SLEDAI score per 1-unit increase	−0.11 (−0.45 to 0.24)	0.55	0.03 (−0.35 to 0.41)	0.88	−0.004 (−0.05 to 0.04)	0.87
SDI score per 1-unit increase	−0.81 (−1.78 to 0.16)	0.10	−0.29 (−1.34 to 0.76)	0.59	−0.04 (−0.16 to 0.08)	0.51
Lupus duration per one-year increase	−0.01 (−0.20 to 0.21)	0.94	0.06 (−0.16 to 0.28)	0.59	0.003 (−0.02 to 0.03)	0.80
Hydroxychloroquine blood levels > 750ng/mL	−0.59 (−2.68 to 1.50)	0.58	0.52 (−1.75 to 2.79)	0.65	−0.05 (−0.30 to 0.21)	0.72
Prednisone dose at visit per 1-mg increase	−0.07 (−0.18 to 0.05)	0.25	−0.03 (−0.16 to 0.10)	0.62	−0.01 (−0.02 to 0.003)	0.13
Immunosuppression	−0.94 (−2.24 to 0.35)	0.15	0.01 (−1.40 to 1.42)	0.99	−0.07 (−0.23 to 0.09)	0.37

* Data are mean change in PROs (95% CI, with $P < 0.05$ indicating significance). Lower scores indicate worse PROs. ADI, Area Deprivation Index; CI, confidence interval; PRO, patient-reported outcome; SDI, Systemic Lupus International Collaborating Clinics Damage Index; SDoH, social drivers of health; SLEDAI, Systemic Lupus Erythematosus Disease Activity Index. Bold font indicates significant *p*-value <0.05.

PRO T score in the physical health domain (95% CI −3.58 to −0.11, P = 0.04; Table 2) and 0.24-point lower PRO T score in the social function domain (95% CI −0.45 to −0.03, P = 0.03; Table 2). However, no significant associations were noted between the presence of economic insecurity and PRO T score in the mental health domain (mean difference −1.32, 95% CI −3.20 to 0.56, P = 0.17; Table 2) after controlling for all factors. Other variables independently associated with lower PROs T scores in adjusted models included unemployment (associated with 3.09 and 0.49 lower PRO T scores in physical and social health domains; Table 2) and being Black or African American (associated with 3.33 and 3.65 lower PRO T scores in physical and mental health domains; Table 2).

Sequential negative impact of increasing economic insecurities. Next, across economic insecurity categories (at least one, at least two, and at least three), a sequential increase in the negative impact of multiple economic insecurities on PROs was noted, particularly in the physical health and social function domains (Supplementary Table 3). To elaborate, patients reporting at least one, at least two, and at least three insecurities had a sequential increase in the mean difference of −2.0, −3.9, and −9.1, respectively, in PRO T scores in the physical health domain (Supplementary Table 3) and −0.3, −0.4, and −1.3, respectively, in the social function domain (Supplementary Table 3). A similar trend showed −1.8, −2.6, and −5.8, respectively, lower T scores in the mental health domain across economic insecurity categories, but these were not statistically significant (P > 0.05; Supplementary Table 3), adjusting for covariables including depression and loneliness (Supplementary Table 3). Our unadjusted models revealed a linear association between increasing economic insecurities and lower patient-reported physical health (P = 0.0005; Supplementary Figure 1A), mental health (P = 0.001; Supplementary Figure 1B), and social function (P = 0.0002; Supplementary Figure 1C).

Subgroup analyses. To further understand interactions between conventional SDoH and economic insecurities, we

Table 3. Variable impact of economic insecurities on patient-reported physical health, mental health, and social function within (1) Medicaid/no insurance, (2) unemployment, (3) Black or African American subgroups*

	Physical health		Mental health		Social function	
Variables at clinic visit	Adjusted mean difference (95% CI)	P value	Adjusted mean difference (95% CI)	P value	Adjusted mean difference (95% CI)	P value
Medicaid/no insurance subgroup (n = 31)						
Economic insecurities per 1-unit increase	**−4.38 (−0.62 to −8.13)**	**0.02**	**−6.47 (−1.94 to −11.00)**	**0.01**	**−0.58 (−0.06 to −1.11)**	**0.03**
Age per 10-year increase	0.09 (−0.17 to 0.34)	0.50	0.26 (−0.05 to 0.57)	0.09	0.01 (−0.03 to 0.04)	0.67
Female	3.42 (−6.21 to 13.00)	0.47	−5.69 (−17.30 to 5.95)	0.32	0.33 (−1.02 to 1.68)	0.62
Black or African American	−3.43 (−10.80 to 3.90)	0.34	−1.11 (−9.97 to 7.74)	0.80	0.28 (−0.75 to 1.30)	0.58
Asian	10.00 (−2.62 to 22.70)	0.12	5.26 (−10 to 20.5)	0.48	0.92 (−0.86 to 2.69)	0.30
Hispanic or Latino	4.91 (−6.97 to 16.80)	0.40	10.40 (−3.96 to 24.70)	0.15	0.33 (−1.33 to 1.99)	0.69
SDI score per 1-unit increase	−3.11 (−7.65 to 1.44)	0.17	−4.59 (−10.10 to 0.90)	0.10	−0.20 (−0.83 to 0.44)	0.53
Unemployed or student subgroup (n = 54)						
Economic insecurities per 1-unit increase	**−2.18 (−0.14 to −4.22)**	**0.04**	−1.85 (−3.95 to 0.25)	0.25	**−0.29 (−0.05 to −0.53)**	**0.02**
Age per 10-year increase	−0.03 (−0.17 to 0.10)	0.64	−0.04 (−0.18 to 0.10)	0.58	−0.01 (−0.02 to 0.01)	0.28
Female	2.01 (−4.36 to 8.37)	0.53	−6.29 (12.80 to −0.26)	0.06	0.01 (−0.73 to 0.75)	0.98
Black or African American	1.08 (−2.94 to 5.10)	0.59	0.53 (−3.61 to 4.66)	0.80	0.16 (−0.31 to 0.63)	0.49
Medicaid insurance/no insurance	−4.08 (−8.34 to 0.19)	0.06	**−5.45 (−1.06 to −9.84)**	**0.02**	**−0.50 (−0.001 to −1.00)**	**0.05**
SDI score per 1-unit increase	−0.55 (−1.93 to 0.84)	0.43	0.12 (−1.30 to 1.55)	0.86	0.06 (−0.10 to 0.23)	0.43
Black or African American subgroup (n = 28)						
Economic insecurities per 1-unit increase	−2.04 (−4.82 to 0.74)	0.15	−1.28 (−4.25 to 1.68)	0.40	−0.08 (−0.42 to 0.25)	0.63
Age per 10-year increase	0.03 (−0.04 to 0.10)	0.35	**0.09 (0.02 to 0.17)**	**0.01**	0.01 (−0.002 to 0.01)	0.13
Female	1.31 (−2.45 to 5.07)	0.49	0.25 (−3.76 to 4.25)	0.90	−0.10 (−0.56 to 0.35)	0.65
Medicaid insurance/no insurance	−1.42 (−5.03 to 2.19)	0.44	−1.99 (−5.84 to 1.86)	0.10	−0.42 (−0.86 to 0.01)	0.054
SDI score per 1-unit increase	**−1.24 (−0.11 to −2.36)**	**0.03**	−0.53 (−1.72 to 0.67)	0.39	**−0.17 (−0.03 to −0.30)**	**0.02**

* Data are adjusted mean differences (95% CI, P < 0.05 indicating significance) in PROs associated with a one-unit increase in economic insecurity within each subgroup (Medicaid/no insurance, unemployment, and Black or African American) compared to no economic insecurity. Adjusted covariates are included. Lower scores indicate worse PROs. CI, confidence interval; PRO, patient-reported outcome; SDI, Systemic Lupus International Collaborating Clinics Damage Index. Bold font indicates significant p-value <0.05.

performed subgroup analyses. In the first subgroup analysis, we only included patients with Medicaid or no insurance. We noted that having economic insecurities in this subgroup had a significant multiplicative negative impact on PROs in all domains (a 4.38-point lower T score in physical health, a 6.47-point lower T score in mental health, and a 0.58-point lower T score in social function; Table 3). This observed change in PRO T scores was higher than the expected additive change of mean PRO T scores (observed −4.38 > expected −2.72 [−1.85 + −0.87]).

The next subgroup included only unemployed participants. Again, having economic insecurities had an adverse impact on PROs in physical and social domains (mean difference in physical health PRO −2.18, 95% CI −4.22 to −0.14, $P = 0.04$; social function PRO −0.29, 95% CI −0.53 to −0.05, $P = 0.02$; Table 3). Finally, in a Black or African American subgroup, economic insecurities were not significantly associated with PROs after controlling for other factors including age, sex, insurance type, and SDI score (Table 3).

Longitudinal impact of assessing and addressing economic insecurities during clinic visits on PROs over time. We used data on PROs at the index clinic visit and then the three- to six-month follow-up visit (after meeting with an SW if indicated based on the presence of one or more economic insecurities). Among 222 total patients, we identified 36 patients with at least one economic insecurity who were referred to an SW or received community support. Among 36 patients, 22 patients (61%) reported resolution of economic barriers at their three- to six-month follow-up. Using an unadjusted pre–post model in this subgroup, PRO T scores significantly improved in the physical health and social function domains by 5.25 (95% CI 1.59–8.91, $P = 0.007$; Supplementary Table 4) and 0.65 (95% CI 0.09–1.21, $P = 0.03$; Supplementary Table 4) compared to baseline PRO scores. Using an adjusted linear regression analysis adjusted for covariables, we noted a significant improvement in PRO T scores in the physical health and social function domains by 4.52 (95% CI 1.35–7.69, $P = 0.005$; Figure 3) and 0.69 (95% CI 0.23–1.14, $P = 0.003$; Figure 3) postintervention compared to the baseline scores. Mean change in PRO T scores in the mental health domain were not statistically significant for both unadjusted (2.41, 95% CI −1.84 to 6.66, $P = 0.25$; Supplementary Table 4) and adjusted (1.12, 95% CI −2.21 to 4.45, $P = 0.51$; Figure 3) models.

DISCUSSION

There is a potential interplay between SDoH and biologic factors, driving poor and heterogeneous SLE outcomes. However, prospective impacts on PROs need to be further elucidated across different populations, particularly identifying and addressing primary nonbiologic (eg, SDoH) drivers of poor outcomes in SLE. Our study directly addresses these gaps by highlighting the following: (1) economic insecurities, such as transportation barriers, financial insecurity, housing instability, and food insecurity, are associated with worse PROs consistent with recent studies[9,13]; (2) a sequential negative impact of increasing economic insecurities on PROs; (3) a systematic and successful approach to screen and address economic insecurities in clinics; and (4) the positive impact of addressing economic insecurities in clinics on PROs, reinforcing their role as modifiable risk factors.

Yelin et al highlighted patients' perspectives on how economic insecurities drove SLE flares and poor physical and mental health.[25] A patient from the study by Yelin et al said, "Money makes it [SLE] worse… I worry about money… and paying my mortgage…"[25,26] Several studies reported that conventional SDoH, such as education, income, poverty, and ADI score, have a negative impact on PROs.[27–29] The CLUES cohort reported that having any economic insecurity was associated with worse PROs regardless of poverty, income, and SLE factors.[9] Economic insecurities, conventional SDoH, and SLE severity vary across geographic regions in the United States.[30] Therefore, the relationship between economic insecurities and PROs needs to be further elucidated in different SLE cohorts. Our findings are timely and important because they validate the negative impact of economic insecurities on PROs in a Midwestern cohort, enhancing the generalizability and reliability of these findings.

Illustrating the sequential burden of economic insecurities on PROs and clinical outcomes can clarify who could benefit from regular screening. Studies in other chronic diseases, such as diabetes, have shown that having multiple (more than two) economic insecurities was linked with a three-fold higher risk of nonadherence and a five-fold higher anxiety rate.[31] Likewise, the CLUES cohort noted significantly lower PRO T scores across physical function, pain, fatigue, sleep, cognitive function, depression, and anxiety domains in patients with multiple economic insecurities.[9] Consistent with these findings, we found that higher counts of economic insecurities (at least one, at least two, and at least three) were associated with progressively worse physical health T scores (−2.0, −3.9, and −9.1, respectively). Similarly, a significant linear association between increasing economic insecurities and lower PROs in all three domains was noted. Cumulatively, our findings support that patients with two or more economic insecurities are at higher risk of poor PROs and may benefit from routine screening for economic insecurities in clinics.

Additionally, quantifying interactions between economic insecurities and conventional SDoH can help explain the complexities driving heterogeneity in clinical outcomes and PROs in SLE. Our subgroup analysis revealed interactions between conventional SDoH and economic insecurities, informing the model shown in Figure 2. For example, having to rely on Medicaid or having no insurance combined with having economic insecurities led to a multiplicative negative impact on PROs, in which the observed PRO change was −4.38 compared to the expected

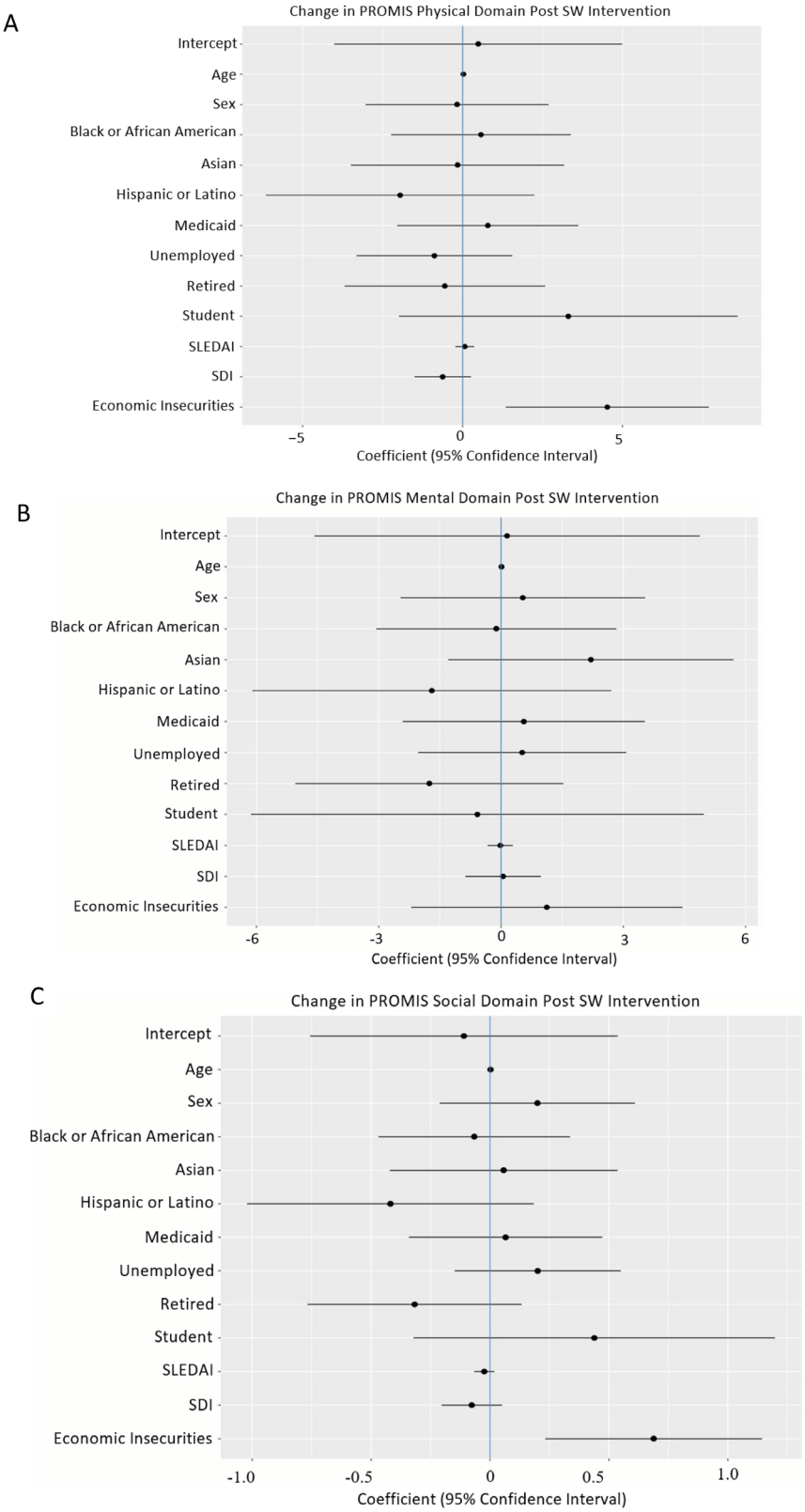

Figure 3. Adjusted mean change in patient-reported (A) physical health, (B) mental health, and (C) social function after social work intervention compared to baseline scores. Points are adjusted means, and error bars are 95% confidence intervals. Data were calculated using pre–post multivariable linear regression analysis adjusting for age, sex, race, ethnicity, Medicaid insurance, employment status, SLEDAI, and SDI. PROMIS, Patient-Reported Outcomes Measurement Information System; SDI, Systemic Lupus International Collaborating Clinics Damage Index; SLEDAI, Systemic Lupus Erythematosus Disease Activity Index; SW, social worker. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80000/abstract.

additive change of –2.72 (–1.85, –0.87). This finding was consistent with our recent meta-analysis highlighting a multiplicative negative impact of having to rely on Medicaid and experiencing care fragmentation or other economic barriers on clinical outcomes in LN.[8] Likewise, a study reported that similar negative interactions among three or more adverse social factors drove a 1.26-fold higher cancer mortality.[32] However, such interactions were not seen in the CLUES cohort. This may reflect differences in Medicaid eligibility and coverage between California versus Wisconsin. Wisconsin has not fully expanded Medicaid, leaving adults in Wisconsin with incomes between 100% and 138% of the federal poverty level without coverage, potentially magnifying the impact of economic insecurities for the uninsured.[33] These differences could explain contrasting interactions between Medicaid insurance and economic insecurities between the two cohorts.

This study provides new evidence on the use of a systematic workflow for SDoH screening and is the first to show that addressing economic insecurities in the clinics has a positive impact on PROs in SLE. To elaborate, both unadjusted and adjusted pre–post analyses showed significant improvements in physical health and social function domains after addressing economic insecurities. These findings are consistent with a systematic review that reported that having an SW in general medicine, pediatrics, or gynecology clinics had an overall positive effect on health outcomes and was less costly because it prevented hospital readmissions.[34] These findings highlight that economic insecurities are modifiable risk factors and integrating routine screening as part of the SLE care plan could be beneficial. Second, the intervention developed for our Collaborative Lupus Clinics is generalizable, yet customizable, across different practices because it leverages a standard, validated EHR tool (Epic SDoH Snapshot) to screen for economic insecurities.[13] Next, it provides workflows to integrate economic insecurities screening in clinics, even in busy practices. Finally, it provides guidance on how to integrate social support or community resources in SLE care plans to offer a comprehensive care plan and improve outcomes across all patients with SLE.

Despite strengths of our study, we do acknowledge limitations. First, data on other conventional SDoH such as poverty or literacy were not available for all patients because these data are not routinely reported in the EHR.[8] Second, because this is a single-center study from the Midwest, it may limit generalizability. Third, because our study leveraged the Epic EHR to assess economic insecurities, this may limit the corroboration of these findings in settings that use different EHR systems. However, the economic insecurities screening questions in Epic follow NAM and CMS guidelines, so similar screening across other institutions and EHRs is likely feasible, with some customization. This will be evaluated in future multicenter pragmatic studies. Finally, given limited sample sizes for patients with higher number of economic insecurities, we used cumulative categories to increase cell sizes for the higher exposure groups. Although this approach improved statistical power, and similar findings were noted with *P* for trend analysis, we acknowledge that larger longitudinal studies are needed to validate our findings.

To summarize, this study is unfolds key information on (1) modifiable high-impact nonbiologic risk factors that should be screened in clinics (eg, transportation barriers, food insecurity, housing instability, financial insecurity) routinely, (2) a systematic method to integrate economic insecurities screening in clinics (eg, Epic SDoH Snapshot), and (3) the positive impact of partnering with an SW to provide a comprehensive care plan that addresses both biologic and nonbiologic factors (eg, SDoH, economic insecurities) on SLE outcomes. These findings lay a roadmap for clinical practice and future clinical research to integrate screening for nonbiologic factors as part of clinical care to advance health in chronic diseases such as SLE.

ACKNOWLEDGMENT

We would like to thank Ms Isabella Hartel and Mr Charlie Dentz for supporting data collection.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Siegel CH, Sammaritano LR. Systemic lupus erythematosus: a review. JAMA 2024;331(17):1480–1491. doi:https://doi.org/10.1001/jama.2024.2315
2. Tektonidou MG, Dasgupta A, Ward MM. Risk of end-stage renal disease in patients with lupus nephritis, 1971-2015: a systematic review and Bayesian meta-analysis. Arthritis Rheumatol 2016;68(6):1432–1441. doi:https://doi.org/10.1002/art.39594
3. Moody WE, Edwards NC, Chue CD, et al. Arterial disease in chronic kidney disease. Heart 2013;99(6):365–372. doi:https://doi.org/10.1136/heartjnl-2012-302818
4. Hermansen ML, Lindhardsen J, Torp-Pedersen C, et al. The risk of cardiovascular morbidity and cardiovascular mortality in systemic lupus erythematosus and lupus nephritis: a Danish nationwide population-based cohort study. Rheumatology (Oxford) 2017;56(5):709–715. doi:https://doi.org/10.1093/rheumatology/kew475
5. Hocaoğlu M, Valenzuela-Almada MO, Dabit JY, et al. Incidence, prevalence, and mortality of lupus nephritis: a population-based study over four decades using the lupus Midwest network. Arthritis Rheumatol 2023;75(4):567–573. doi:https://doi.org/10.1002/art.42375
6. Utset TO, Baskaran A, Segal BM, et al. Work disability, lost productivity and associated risk factors in patients diagnosed with systemic lupus erythematosus. Lupus Sci Med 2015;2(1):e000058. doi:https://doi.org/10.1136/lupus-2014-000058

7. Ramos PS. Integrating genetic and social factors to understand health disparities in lupus. Curr Opin Rheumatol 2021;33(6):598–604. doi: https://doi.org/10.1097/BOR.0000000000000840
8. Garg S, Sweet N, Boderman B, et al. Multiplicative impact of adverse social determinants of health on outcomes in lupus nephritis: a meta-analysis and systematic review. Arthritis Care Res (Hoboken) 2024; 76(9):1232–1245. doi:https://doi.org/10.1002/acr.25359
9. Sandoval-Heglund D, Roberts E, Park J, et al. Economic insecurities and patient-reported outcomes in patients with systemic lupus erythematosus in the USA: a cross-sectional analysis of data from the California Lupus Epidemiology Study. Lancet Rheumatol 2024;6(2): e105–e114. doi:https://doi.org/10.1016/S2665-9913(23)00296-5
10. Borchers AT, Naguwa SM, Shoenfeld Y, et al. The geoepidemiology of systemic lupus erythematosus. Autoimmun Rev 2010;9(5):A277–A287. doi:https://doi.org/10.1016/j.autrev.2009.12.008
11. Schletzbaum M, Powell WR, Garg S, et al. Receipt of rheumatology care and lupus-specific labs among young adults with systemic lupus erythematosus: a US Medicare retention in care cohort study. Lupus 2024; 33(8):804–815. doi:https://doi.org/10.1177/09612033241247905
12. Garg S, Singh T, Panzer SE, et al. Multidisciplinary lupus nephritis clinic reduces time to renal biopsy and improves care quality. ACR Open Rheumatol 2022;4(7):581–586. doi:https://doi.org/10.1002/acr2.11435
13. Chandler A, Hamid M, Jiao Z, et al. Patient-reported social risk factor screening among rheumatology outpatients. ACR Open Rheumatol 2025;7(3):e70018. doi:https://doi.org/10.1002/acr2.70018
14. Chandler MT, Cai T, Santacroce L, et al. Classifying individuals with rheumatic conditions as financially insecure using electronic health record data and natural language processing: algorithm derivation and validation. ACR Open Rheumatol 2024;6(8):481–488. doi: https://doi.org/10.1002/acr2.11675
15. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1997;40(9):1725. doi:https://doi.org/10.1002/art.1780400928
16. Petri M, Orbai AM, Alarcón GS, et al. Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus. Arthritis Rheum 2012;64(8): 2677–2686. doi:https://doi.org/10.1002/art.34473
17. Aringer M, Costenbader K, Daikh D, et al. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. Arthritis Rheumatol 2019;71(9):1400–1412. doi:https://doi.org/10.1002/art.40930
18. Hays RD, Bjorner JB, Revicki DA, et al. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. Qual Life Res 2009;18(7):873–880. doi:https://doi.org/10.1007/s11136-009-9496-9
19. Gladman DD, Ibañez D, Urowitz MB. Systemic Lupus Erythematosus Disease Activity Index 2000. J Rheumatol 2002;29(2):288–291.
20. Gladman DD, Urowitz MB, Rahman P, et al. Accrual of organ damage over time in patients with systemic lupus erythematosus. J Rheumatol 2003;30(9):1955–1959.
21. Garg S, Astor BC, Saric C, et al. Therapeutic hydroxychloroquine blood levels are associated with fewer hospitalizations and possible reduction of health disparities in lupus. Arthritis Care Res (Hoboken) 2024;76(12):1606–1616. doi:https://doi.org/10.1002/acr.25422
22. Garg S, Unnithan R, Hansen KE, et al. Clinical significance of monitoring hydroxychloroquine levels in patients with systemic lupus erythematosus: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2021;73(5):707–716. doi:https://doi.org/10.1002/acr.24155
23. Garg S, Chewning B, Hutson P, et al. Reference range of hydroxychloroquine blood levels that can reduce odds of active lupus and prevent flares. Arthritis Care Res (Hoboken) 2024;76(2):241–250. doi:https://doi.org/10.1002/acr.25228
24. Global Health Scoring Manual. Version 1.0. PROMIS; 2021. Accessed June 05, 2024. https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Global_Health_Scoring_Manual_22June2023.pdf
25. Yelin E, Trupin L, Bunde J, et al. Poverty, neighborhoods, persistent stress, and systemic lupus erythematosus outcomes: a qualitative study of the patients' perspective. Arthritis Care Res (Hoboken) 2019;71(3):398–405. doi:https://doi.org/10.1002/acr.23599
26. Watson B, Osberg L. Healing and/or breaking? The mental health implications of repeated economic insecurity. Soc Sci Med 2017; 188:119–127. doi:https://doi.org/10.1016/j.socscimed.2017.06.042
27. Hoge C, Bowling CB, Lim SS, et al. Association of poverty income ratio with physical functioning in a cohort of patients with systemic lupus erythematosus. J Rheumatol 2020;47(7):983–990. doi:https://doi.org/10.3899/jrheum.190991
28. Trupin L, Tonner MC, Yazdany J, et al. The role of neighborhood and individual socioeconomic status in outcomes of systemic lupus erythematosus. J Rheumatol 2008;35(9):1782–1788.
29. Williams JN, Drenkard C, Lim SS. The impact of social determinants of health on the presentation, management and outcomes of systemic lupus erythematosus. Rheumatology (Oxford) 2023;62(suppl 1):i10–i14. doi:https://doi.org/10.1093/rheumatology/keac613
30. Walsh SJ, DeChello LM. Geographical variation in mortality from systemic lupus erythematosus in the United States. Lupus 2001;10(9): 637–646. doi:https://doi.org/10.1191/096120301682430230
31. Leung CW, Heisler M, Patel MR. Multiple social risk factors are adversely associated with diabetes management and psychosocial outcomes among adults with diabetes. Prev Med Rep 2022;29: 101957. doi:https://doi.org/10.1016/j.pmedr.2022.101957
32. Pinheiro LC, Reshetnyak E, Akinyemiju T, et al. Social determinants of health and cancer mortality in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) cohort study. Cancer 2022;128(1): 122–130. doi:https://doi.org/10.1002/cncr.33894
33. State-by-state Medicaid expansion guide. HealthInsurance.org. 2024. Accessed December 24, 2025. https://www.healthinsurance.org/glossary/medicaid-expansion/2025
34. Steketee G, Ross AM, Wachman MK. Health outcomes and costs of social work services: a systematic review. Am J Public Health 2017; 107(S3):S256–S266. doi:https://doi.org/10.2105/AJPH.2017.304004

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1188–1194
DOI 10.1002/acr.70042
Published 2026. This article is a U.S. Government work and is in the public domain in the USA.

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Systemic Lupus Erythematosus Mortality Among Decedents Aged ≥15 Years—United States, 2018–2023

Danielle Dawson, MPH,[1,2] Kurt J. Greenlund, PhD,[1] and Kamil E. Barbour, PhD[1]

Objective. Systemic lupus erythematosus (SLE) is a chronic autoimmune condition that can lead to death. To examine SLE as an underlying and contributing cause of death, the Centers for Disease Control and Prevention (CDC) analyzed 2018 to 2023 mortality data for persons aged ≥15 years overall and by age, sex, race and ethnicity, and region.

Methods. Death certificate data for persons aged ≥15 years with any mention of SLE (*International Classification of Diseases, Tenth Revision* [ICD-10] code M32) were analyzed using the CDC Wide-ranging Online Data for Epidemiologic Research (WONDER) system. We calculated age-adjusted death rates and assessed patterns by sex, age, race and ethnicity, and region. Underlying and contributing causes of death were evaluated using ranked cause-of-death lists and ICD-10 subchapters.

Results. During 2018 to 2023, 14,936 deaths had any mention of SLE listed on the death certificates. Of these deaths, 6,414 (42.9%) listed SLE as the underlying cause. The age-adjusted SLE mortality rate per million population was greater among females (5.97) than males (1.16), non-Hispanic African American persons (10.70) than persons of other non-Hispanic racial groups (range 2.46–5.62), Hispanic persons (3.98) than non-Hispanic persons (3.59), and people in the South (4.37) than people in other regions. When SLE was listed as a contributing cause of death, the leading underlying causes were heart disease (0.93), cancer (0.56), and COVID-19 (0.51). The overall age-adjusted SLE mortality rates were significantly higher in 2020 and 2021 than in all other study years, indicating the likely impact of the COVID-19 pandemic on SLE mortality.

Conclusion. Overall management of SLE, comorbidities, and infections in patients with SLE, as well as interventions targeting groups (eg, African American persons) disproportionately impacted by SLE, may reduce overall SLE mortality.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a chronic autoimmune condition that impacts over 200,000 persons in the United States.[1] It affects many organs such as the skin, heart, brain, joints, and kidneys.[2] SLE is much more common in females than males, and females of childbearing age have the highest risk of developing SLE.[1,2] Patients with SLE have a three-fold higher risk of dying than the general population.[3] Globally, cardiovascular disease, infections, and renal disease (eg, lupus nephritis and end-stage renal disease) contribute to higher mortality among those with SLE.[3] Several studies have examined SLE as an underlying cause of death in the United States.[4–6] However, to our knowledge, few US-based studies have examined causes of death for which SLE is listed as the underlying versus a contributing cause.[5] These results can inform management and treatment for patients with SLE. Moreover, little is known nationally about how the COVID-19 pandemic may have impacted SLE mortality. To provide the most up-to-date national estimates of SLE mortality, we assessed SLE deaths in the United States from 2018 to 2023 overall and by age, sex, race and ethnicity, and region. Furthermore, to address the aforementioned gaps in the literature,

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

[1]Centers for Disease Control and Prevention, Atlanta, Georgia; [2]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.70042).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.70042.

Address correspondence via email to Kamil E. Barbour, PhD, at kamil.barbour@nih.gov.

Submitted for publication August 7, 2025; accepted in revised form December 15, 2025.

SIGNIFICANCE & INNOVATIONS

- From 2018 to 2023, there were 14,936 deaths among persons aged ≥15 years with any mention of SLE on their death certificates (6,414 as underlying and 8,522 as contributing causes of death).
- Non-Hispanic African American females had the highest age-adjusted SLE underlying mortality rate (17.6 per million).
- Cardiovascular disease, cancer, and COVID-19 were the leading underlying causes of death for which SLE was not listed as the primary cause.
- Age-adjusted overall SLE mortality rates were significantly higher in 2020 and 2021 than in all other study years, likely indicating the impact of the COVID-19 pandemic on SLE mortality, with an increase in the number of SLE deaths and a shift to listing COVID-19 as the underlying cause among SLE deaths.
- From 2020 to 2023, COVID-19 was more likely to be listed as a cause of death among those for whom SLE was listed on their death certificates (10.8%) compared to the overall population (8.9%), indicating greater vulnerability to COVID-19 mortality in the population with SLE.
- Other underlying causes of death of importance in SLE beyond the 15 leading causes of death included coagulation defects and scleroderma.
- Prevention strategies targeting at-risk populations may reduce disparities in SLE-related mortality.

we assessed the major underlying causes for which SLE was listed on the death certificates but not as the primary cause of death using two complimentary approaches for the 15 leading causes of death (including COVID-19) as well as other underlying causes of death. Finally, we examined SLE mortality rates (overall, underlying cause, and contributing cause) by year to understand the potential impact of the COVID-19 pandemic on mortality.

MATERIALS AND METHODS

The Centers for Disease Control and Prevention (CDC) analyzed 2018 to 2023 mortality data for persons aged ≥15 years with any mention of SLE on their death certificates using the CDC Wide-ranging Online Data for Epidemiologic Research (WONDER; wonder.cdc.gov) system. CDC WONDER uses death certificate data to provide underlying and multiple causes of death. The underlying cause of death was defined as "the disease or injury that initiated the events resulting in death." Contributing cause of death was defined as "any cause of death that is neither the direct, intervening, originating antecedent nor underlying." SLE deaths were identified based on any mention of SLE, ie, presence of the *International Classification of Diseases, Tenth Revision* (ICD-10) code M32 on the death records. The analysis focused on deaths among those ≥15 years to protect privacy and to avoid unreliably small numbers for younger age groups. CDC WONDER identifies <20 deaths within a group as statistically unreliable, and estimates are suppressed. We examined rates for SLE as the underlying cause of death by demographic characteristics as well as leading underlying causes of death for which SLE was recorded anywhere on the death certificates but not listed as the underlying cause (referred to here as "contributing" cause).

Age-adjusted death rates (standardized to the projected 2000 US population based on National Center for Health Statistics [NCHS] guidance) per 1,000,000 with 95% confidence intervals (CIs) were calculated for SLE as the underlying cause to compare groups that have different age distributions.[7] A conservative approach (nonoverlapping CIs) was used to assess statistical significance when comparing age-adjusted mortality rates between categories.[8] Age-adjusted death rates were also calculated for the major underlying causes of death for which SLE was listed as a contributing cause. We first identified the 15 leading underlying causes of death based on the NCHS list of 52 ranked causes of death.[9] Because the list of selected ranked causes may not capture all causes of death that may be important in SLE, we also examined causes of death based on ICD-10 subchapters to identify other cause of death categories that may have contributed significantly to SLE mortality, removing overlap of deaths with the leading cause list. We report only those ICD-10 subchapter causes of death with a considerable number of deaths that would be comparable to the 15 ranked causes. To address the issue of low positive predictive value (leading to an increased probability of false positive results) for SLE when using ICD-10 codes, we performed a sensitivity analysis excluding participants with both SLE and scleroderma (on their death certificate) to reduce the likelihood that deaths from scleroderma were not attributed to SLE. This activity was reviewed by the CDC, deemed not to be research, and conducted consistent with applicable federal law and CDC policy. Thus, institutional review board approval was not required for this study using publicly available data.

RESULTS

During 2018 to 2023, there were 14,936 deaths among persons aged ≥15 years with any mention of SLE. Annual deaths ranged from 2,207 in 2018 to 2,918 in 2021, with 2,291 deaths in 2023. Overall age-adjusted rates per million were significantly higher in 2020 (9.36 [95% CI 9.00–9.71]) and 2021 (9.71 [95% CI 9.35–10.07]) compared with 2018 (7.64 [95% CI 7.31–7.96]), 2019 (7.83 [95% CI 7.51–8.16]), 2022 (7.79 [95% CI 7.47–8.11]), and 2023 (7.30 [95% CI 7.00–7.61]) (Figure 1). From 2018 to 2021, the age-adjusted mortality rate for SLE as the underlying cause remained steady with no significant changes: 2018 (4.10 [95% CI 3.86–4.34]), 2019 (4.05 [95% CI 3.82–4.29]), 2020 (4.26 [95% CI 4.01–4.50]), 2021(4.05 [95% CI 3.83–4.29]); however, the age-adjusted rate for SLE listed as a

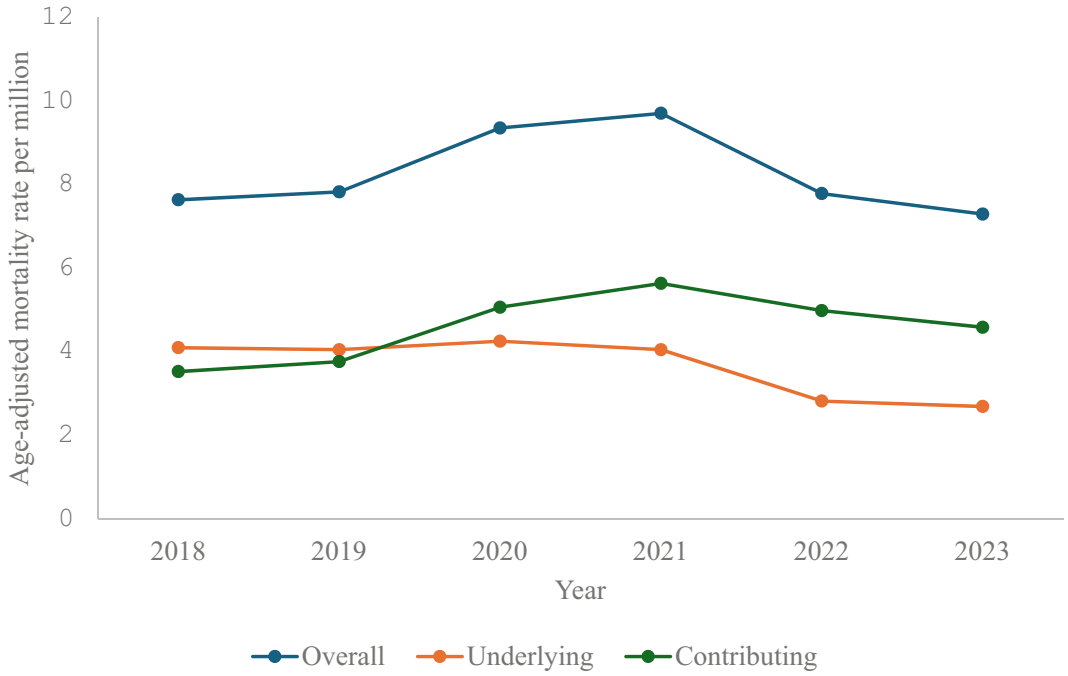

Figure 1. Systemic lupus erythematosus age-adjusted mortality rate overall, underlying cause of death, and contributing cause of death (per million) by year.

contributing cause increased significantly from 2018 (3.53 [95% CI 3.31–3.75]) and 2019 (3.77 [95% CI 3.55–3.99]) to 2020 (5.07 [95% CI 4.81–5.33]) and 2021 (5.64 [95% CI 5.37–5.91]) (Figure 1), indicating a likely impact of the COVID-19 pandemic on mortality rates among those with SLE. The overall SLE mortality rates over time did not differ much from the overall trends when stratifying by age, sex, and race and ethnicity, with the exception of those aged 15 to 44 years, males, and Asian non-Hispanic people, in which prepandemic (2018 and 2019) and pandemic rates in 2020 and 2021 were statistically similar (Supplementary Figures 1–3).

After excluding participants with both SLE and scleroderma on their death certificates (n = 944), there were 13,922 deaths (5,516 underlying and 8,476 contributing) among persons aged ≥15 years with any mention of SLE (Supplementary Figure 4). The overall age-adjusted SLE mortality trend over time followed the original sample, with age-adjusted rates per million peaking in 2020 (8.76 [95% CI 8.42–9.12]) and 2021 (9.17 [95% CI 8.82–9.52]) compared with other years. The trends for underlying and contributing age-adjusted SLE mortality were also similar for the sensitivity analysis.

SLE was listed as the underlying cause for 6,414 of the 14,936 deaths (42.9%) in 2018 to 2023 (Table 1). The age-adjusted mortality rate per million population for SLE as the underlying cause of death was greater among females (5.97 [95% CI 5.81–6.14]) than males (1.16 [95% CI 1.08–1.23]), increased with age, and was highest among non-Hispanic African Americans (10.70 [95% CI 10.23–11.16]), among Hispanic versus non-Hispanic persons (3.98 [95% CI 3.72–4.24] vs 3.59 [95% CI 3.49–3.69]), and in the South (4.37 [95% CI 4.21–4.53]) versus other regions. In addition, the age-adjusted rate was higher for non-Hispanic African American females (17.58 [95% CI 16.77–18.39]) compared with females of other non-Hispanic groups. The age-adjusted rates were also higher in non-Hispanic African American males (2.68 [95% CI 2.34–3.02]) than non-Hispanic White males (0.86 [95% CI 0.78–0.93]) and non-Hispanic Asian males (0.77 [95% CI 0.53–1.07]). The age-adjusted rates were higher in Hispanic females (6.63 [95% CI 6.17–7.09]) than non-Hispanic females (5.87 [95% CI 5.69–6.05]); however, there was no difference between Hispanic and non-Hispanic males.

During 2018 to 2023, there were 8,522 deaths with SLE listed as a contributing cause of death. Similar mortality rate patterns by demographics were observed for SLE as a contributing cause of death as for SLE as an underlying cause (Supplementary Table 1).

The 15 leading underlying causes of death accounted for 6,204 of the 8,522 deaths with SLE listed as a contributing cause. The most common underlying cause of death was heart disease (1,750 deaths; age-adjusted rate 0.93 [95% CI 0.89–0.98]), followed by cancer (1,142 deaths; age-adjusted rate 0.56 [95% CI 0.52–0.59]) and COVID-19 (952 deaths; age-adjusted rate 0.51 [95% CI 0.47–0.54]) (Table 2). Mortality rates were substantially lower for the other 12 leading causes of death.

Further, from 2020 to 2023, COVID-19 was more likely to be listed as any (underlying or contributing) cause of death among those for whom SLE was listed on their death certificates

Table 1. SLE as the underlying cause of death among persons aged ≥15 years, overall and by age, sex, race, ethnicity, and geographic region—United States, 2018–2023*

Characteristics	Deaths[a]	Age-adjusted mortality rate[b] per 1,000,000 population (95% CI)
Overall	6,414	3.67 (3.57–3.76)
Age, y[c]		
15–44	1,596	2.02 (1.92–2.12)
45–64	2,259	4.53 (4.35–4.72)
≥65	2,559	7.64 (7.34–7.93)
Sex		
Female	5,440	5.97 (5.81–6.14)
Male	974	1.16 (1.08–1.23)
Race, ethnicity, and sex[d]		
American Indian or Alaska Native, non-Hispanic	64	5.62 (4.30–7.22)
Female	NR[e]	NR[e]
Male	NR[e]	NR[e]
Asian, non-Hispanic	249	2.47 (2.16–2.78)
Female	215	3.96 (3.43–4.50)
Male	34	0.77 (0.53–1.07)
Black or African American, non-Hispanic	2,142	10.70 (10.23–11.16)
Female	1,889	17.58 (16.77–18.39)
Male	253	2.68 (2.34–3.02)
Black or African American, Hispanic	23	1.99 (1.23–3.04)
Female	NR[e]	NR[e]
Male	NR[e]	NR[e]
Native Hawaiian or other Pacific Islander, non-Hispanic	NR[e]	NR[e]
Female	NR[e]	NR[e]
Male	NR[e]	NR[e]
White, non-Hispanic	2,892	2.24 (2.15–2.32)
Female	2,369	3.50 (3.35–3.65)
Male	523	0.86 (0.78–0.93)
White, Hispanic	934	4.24 (3.95–4.52)
Female	801	6.55 (6.09–7.00)
Male	133	1.07 (0.89–1.25)
More than one race, non-Hispanic	57	2.46 (1.84–3.23)
Female	NR[e]	NR[e]
Male	NR[e]	NR[e]
More than one race, Hispanic	NR[e]	NR[e]
Female	NR[e]	NR[e]
Male	NR[e]	NR[e]
Ethnicity and sex[f]		
Hispanic (any race)	980	3.98 (3.72–4.24)
Female	839	6.63 (6.17–7.09)
Male	141	1.11 (0.92–1.31)
Non-Hispanic (any race)	5,421	3.59 (3.49–3.69)
Female	4,590	5.87 (5.69–6.05)
Male	831	1.16 (1.08–1.24)
Region		
Northeast	874	2.86 (2.66–3.05)
Midwest	1,143	3.08 (2.90–3.27)
South	2,914	4.37 (4.21–4.53)
West	1,483	3.61 (3.42–3.80)

* CI, confidence interval; NR, not reliable; SLE, systemic lupus erythematosus.
[a] Number of deaths and population are for the full 2018–2023 period.
[b] Age-adjusted mortality rates were calculated using the projected 2000 US standard population.
[c] Crude (unadjusted) mortality rates for age category are shown.
[d] The Hispanic population for American Indian, Asian, Native Hawaiian, or other Pacific Islander groups was suppressed for this category. Moreover, estimates for race and ethnicity may be subject to misclassification. See https://www.cdc.gov/nchs/data/series/sr_02/sr02_172.pdf.
[e] Death rates are considered unreliable in cases in which the rates for one sex are suppressed because this requires the suppression of the entire category. Numbers and estimates are suppressed according to National Center for Health Statistics reporting guidelines (https://wonder.cdc.gov/wonder/help/faq.html).
[f] There were 13 decedents whose Hispanic origin was not stated.

Table 2. Underlying causes of death in the United States among people aged 15 years and older with SLE listed as a contributing cause of death—United States, 2018–2023 (N = 7,527)*

Underlying causes of death (ICD-10 codes)	Deaths[a] (leading causes), n (%)	Deaths[b] (ICD-10 subchapter), n (%)	Age-adjusted mortality rate[c] per 1,000,000 (95% CI)
Heart disease (I00–I09, I11, I13, I20–I51)	1,750 (23.2)		0.93 (0.89–0.98)
Cancer (C00–C97)	1,142 (15.2)		0.56 (0.52–0.59)
COVID-19 (U07.1)	952 (12.6)		0.51 (0.47–0.54)
Chronic lower respiratory diseases (J40–J47)	388 (5.2)		0.19 (0.17–0.21)
Cerebrovascular disease (I60–I69)	379 (5.0)		0.21 (0.18–0.23)
Diabetes mellitus (E10–E14)	322 (4.3)		0.17 (0.18–0.23)
Accidents (V01–X59, Y85–Y86)	224 (3.0)		0.10 (0.09–0.12)
Chronic liver disease and cirrhosis (K70, K73–K74)	214 (2.8)		0.11 (0.10–0.13)
Septicemia (A40–A41)	205 (2.7)		0.11 (0.09–0.13)
Nephritis, nephrotic syndrome, and nephrosis (N00–N07, N17–N19, N25–N27)	168 (2.2)		0.10 (0.08–0.11)
Organic, including symptomatic, mental disorders (F01–F09)		144 (1.9)	0.07 (0.06–0.08)
Essential hypertension and hypertensive renal disease (I10, I12, I15)	138 (1.8)		0.07 (0.05–0.08)
Alzheimer disease (G30)	137 (1.8)		0.07 (0.06–0.09)
Coagulation defects, purpura, and other hemorrhagic conditions (D65–D69)		135 (1.8)	0.07 (0.05–0.08)
Other diseases of intestines (K55–K63)		119 (1.6)	0.06 (0.05–0.08)
Diseases of arteries, arterioles, and capillaries (I70–I78)		114 (1.5)	0.05 (0.04–0.07)
Other respiratory diseases principally affecting the interstitium (J80–J84)		110 (1.5)	0.04 (0.03–0.05)
Systemic connective tissue disorders excluding SLE (M32, M30–M35)[d]		102 (1.4)	0.05 (0.04–0.06)
Metabolic disorders (E70–E88)		99 (1.3)	0.03 (0.03–0.04)
Diseases of liver (K70–K76)[e]		94 (1.2)	0.04 (0.03–0.05)
Other degenerative diseases of the nervous system (G30, G31)[f]		73 (1.0)	0.04 (0.03–0.05)
Influenza and pneumonia (J09–J18)	69 (0.9)		0.03 (0.02–0.04)
Nutritional deficiencies (E40–E64)	65 (0.9)		0.02 (0.01–0.03)
Arthropathies (M00–M25)		58 (0.8)	0.03 (0.02–0.04)
Intestinal infectious diseases (A00–A09)		58 (0.8)	0.02 (0.02–0.03)
Certain disorders involving the immune mechanism (D80–D89)		55 (0.7)	0.03 (0.02–0.04)
Other disorders of the skin and subcutaneous tissue (L80–L98)		55 (0.7)	0.01 (0.01–0.02)
Other diseases of urinary system (N30–N39)		54 (0.7)	0.02 (0.01–0.03)
Other diseases of blood and blood-forming organs (D70–D76)		53 (0.7)	0.00 (0.00–0.00)
In situ neoplasms, benign neoplasms, and neoplasms of uncertain or unknown behavior (D00–D48)	51 (0.7)		0.02 (0.01–0.02)

* CI, confidence interval; ICD-10, *International Classification of Diseases, Tenth Revision*; SLE, systemic lupus erythematosus.
[a] Leading rankable causes of death categorized by the National Center for Health Statistics (https://www.cdc.gov/nchs/data/nvsr/nvsr73/nvsr73-04.pdf).
[b] Nonranked causes of death based on ICD-10 subchapters.
[c] Age-adjusted mortality rates were calculated using the 2000 US standard population.
[d] ICD-10 code M32 (SLE) in this subchapter is excluded.
[e] ICD-10 codes K70 and K73–K74 in this subchapter are reported as ranked causes of death (chronic liver disease and cirrhosis) and are excluded here.
[f] ICD-10 code G30 in this subchapter is reported as a ranked cause of death (Alzheimer disease) and is excluded here.

(10.8%) than the overall population (8.9%). In addition, the age-adjusted mortality rate in the pandemic years (2020–2023) for the leading underlying causes of death was slightly higher than that in the prepandemic years (2019–2020), but the rank among the leading causes of death was consistent, with the exception of COVID-19 becoming a leading cause of death during the pandemic.

Other underlying causes of death (Table 2)—grouped by ICD-10 subchapters—contributed to as many deaths as some of the five lowest ranked leading causes of death. For instance, there were 135 deaths with underlying causes of death related to coagulation defects, purpura, and other hemorrhagic conditions (D65–D69), with the majority (79%) of these listed as "other coagulation defects" (D68). Moreover, the categories M30 to

M35 (excluding M32) for systemic connective tissue disorders included 102 deaths in which SLE was a contributing factor, of which 46 (45%) were listed as systemic sclerosis (scleroderma; M34).

DISCUSSION

In summary, from 2018 to 2023, there were 14,936 deaths among persons aged ≥15 years with any mention of SLE (6,414 as underlying and 8,522 as contributing causes of death). Non-Hispanic African American females had the highest age-adjusted SLE underlying mortality rate by far (a rate of 17.6 per million). Cardiovascular disease, cancer, and COVID-19 were easily the leading underlying causes of death when SLE was listed as a contributing cause. Overall, age-adjusted SLE mortality rates were significantly higher in 2020 and 2021 compared with all other study years, indicating the likely potential impact of the COVID-19 pandemic on SLE mortality. This can be seen by the increase in overall SLE deaths and death rates and COVID-19 becoming a leading underlying cause of death among SLE-associated deaths. However, these pandemic-related trends were less prominent among those aged 15 to 44 years, males, and Asian non-Hispanic people, in whom prepandemic (2018 and 2019) and pandemic rates in 2020 and 2021 were statistically similar. From 2020 to 2023, COVID-19 was more likely to be listed as a cause of death among those for whom SLE was listed on their death certificates (10.8%) than the overall population (8.9%), indicating a greater vulnerability to COVID-19 mortality in the population with SLE. Finally, a more detailed assessment of mortality also identified other underlying causes of death of importance in SLE beyond the 15 leading causes of death, including coagulation defects and scleroderma.

These results support prior studies that identified higher SLE mortality among older adults, in women compared with men, for African American people compared with people of other racial and ethnic groups, and in the South compared with other regions of the United States.[3–7,10] The SLE mortality rate increased exponentially with greater age. This could be related to multiple factors, including a late diagnosis of SLE at an older age, which may result in more severe disease and, thus, a lower survival probability; furthermore, older adults are more likely to have comorbidities that can complicate the treatment of SLE, increasing the risk of death. Medical factors, including severity of SLE and certain comorbid health conditions, that increase risk of death could explain some of the differences in mortality.[10] Nonmedical factors, including poverty, unemployment, food insecurity, housing affordability, crime, stress, racial segregation, and discrimination, are factors for poor outcomes in SLE resulting in health disparities among certain racial and ethnic groups.[10,11]

The leading underlying cause of death for which SLE was listed as a contributing cause was heart disease, but mortality from cancer and COVID-19 was also substantial. This was observed with an increase in overall SLE death rates during the pandemic years and with SLE more likely to be listed as a contributing cause than as the underlying cause. Previous studies have observed increased cardiovascular risk for people with lupus.[3,10] A prior study also observed that patients with SLE have an increased risk of severe outcomes (mechanical ventilation and hospitalization) and an over two-fold risk of mortality from COVID-19 compared with those without SLE.[12] In our study, COVID-19 was identified as the third leading cause of death for which SLE was a contributing cause. Moreover, the prominence of COVID-19 in 2020 and 2021 also coincided with the highest age-adjusted overall SLE mortality rates in this study. Thus, in addition to traditional risk factors for SLE mortality (eg, cardiovascular disease and cancer), health professionals should consider COVID-19 as a major risk factor for mortality in patients with SLE due to the impact of the virus on immunocompromised individuals.

A more in-depth assessment of mortality also identified other underlying causes of death of importance in SLE. For example, coagulation defects showed a substantial number of deaths; prior research identified SLE as a stand-alone risk factor for the development of venous and arterial thrombotic events.[13] A cohort study showed that the prevalence of co-occurrence of SLE among participants with scleroderma was high (6.8%), supporting our observation of scleroderma as an underlying cause of death with SLE.[14]

Beyond treatment and management of SLE, prevention and management of comorbidities or co-occurring conditions may lower mortality risk in people with SLE. For example, cardiovascular disease is a well-recognized complication of SLE, and patients are at a two- to three-fold higher risk of cardiovascular disease mortality than the general population.[3,10] The Chronic Disease Self-Management Program has been shown to improve outcomes and self-efficacy in Black women with SLE, a group at high risk. Preventive measures focusing on African American women might also focus on health disparities to understand how they are disproportionately affected by lupus. Moreover, because COVID-19 has become endemic, COVID-19 vaccination, good hygiene practices, proper mask wearing, and avoiding close contact with infected individuals could lower the risk of severe illnesses and COVID-19 transmission.[15]

There are at least five limitations to this study. First, mortality rates could not be estimated for population groups with fewer than 20 deaths because estimates were unreliable and may impact privacy. This impacts our ability to report mortality rates and limits our understanding of the impact of SLE on mortality in these unique populations. Second, racial identity and ethnicity were based on information reported on death certificates and may not accurately reflect the decedents' actual race or ethnicity. This information bias potentially can lead to misclassification of race and ethnicity in our study population. Nonetheless, the race and ethnicity results were similar to what was observed in prior

reports.[4–6] Third, because the rate of patients with underdiagnosed or misdiagnosed SLE remains substantial, the SLE mortality rate may be underestimated if SLE cases were not captured. Fourth, underestimation may have occurred if certifiers failed to specify "systemic" when reporting lupus as a cause of death. In such cases, coding defaults to L93.0 (discoid lupus erythematosus) instead of M32 (SLE), leading to potential misclassification. Finally, the SLE underlying mortality rate may be underestimated because our analysis does not account for competing risks of mortality, an important consideration, particularly among patients with SLE who may have multiple comorbidities. The presence of other comorbidities may influence the probability of death from SLE-related causes.

In conclusion, management and treatment of SLE as well as preventive measures that address comorbid conditions among people with SLE may lower the risk of mortality for people with SLE.

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 Barbour 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. Izmirly PM, Parton H, Wang L, et al. Prevalence of systemic lupus erythematosus in the United States: estimates from a meta-analysis of the Centers for Disease Control and Prevention National Lupus Registries. Arthritis Rheumatol 2021;73(6):991–996. doi:https://doi.org/10.1002/art.41632
2. Weckerle CE, Niewold TB. The unexplained female predominance of systemic lupus erythematosus: clues from genetic and cytokine studies. Clin Rev Allergy Immunol 2011;40(1):42–49. doi:https://doi.org/10.1007/s12016-009-8192-4
3. Lee YH, Song GG. Mortality in patients with systemic lupus erythematosus: a meta-analysis of overall and cause-specific effects. Lupus 2024;33(9):929–937. doi:https://doi.org/10.1177/09612033241257134
4. Daoud A, Dweik L, Pamuk O. Temporal trends and demographic insights into mortality from systemic lupus erythematosus, 1999-2020. Arthritis Care Res (Hoboken) 2025;77(8):988–997. doi:https://doi.org/10.1002/acr.25509
5. Yen EY, Shaheen M, Woo JMP, et al. 46-year trends in systemic lupus erythematosus mortality in the United States, 1968 to 2013: a nationwide population-based study. Ann Intern Med 2017;167(11):777–785. doi:https://doi.org/10.7326/M17-0102
6. Patel S, Bettag J, Baral N, et al. Trends in systemic lupus erythematous mortality in the United States, 1999-2022. ACR Open Rheumatol 2025;7(6):e70057. doi:https://doi.org/10.1002/acr2.70057
7. Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected U.S. population. Healthy People 2000 Stat Notes 2001;(20):1–9.
8. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. Am Stat 2001;55(3):182–186. doi:https://doi.org/10.1198/000313001317097960
9. Curtin SC, Tejada-Vera B, Bastian BA. Deaths: leading causes for 2021. Natl Vital Stat Rep 2024;73(4). https://doi.org/10.15620/cdc/147882
10. Taylor T, Anastasiou C, Ja C, et al. Causes of death among individuals with systemic lupus erythematosus by race and ethnicity: a population-based study. Arthritis Care Res (Hoboken) 2023;75(1):61–68. doi:https://doi.org/10.1002/acr.24988
11. Sandoval-Heglund D, Roberts E, Park J, et al. Economic insecurities and patient-reported outcomes in patients with systemic lupus erythematosus in the USA: a cross-sectional analysis of data from the California Lupus Epidemiology Study. Lancet Rheumatol 2024;6(2):e105–e114. doi:https://doi.org/10.1016/S2665-9913(23)00296-5
12. Bruera S, Lei X, Zhao H, et al. Risks of mortality and severe coronavirus disease 19 (COVID-19) outcomes in patients with or without systemic lupus erythematosus. Lupus Sci Med 2023;10(1):e000750. doi:https://doi.org/10.1136/lupus-2022-000750
13. Bazzan M, Vaccarino A, Marletto F. Systemic lupus erythematosus and thrombosis. Thromb J 2015;13(1):16. doi:https://doi.org/10.1186/s12959-015-0043-3
14. Alharbi S, Ahmad Z, Bookman AA, et al. Epidemiology and survival of systemic sclerosis-systemic lupus erythematosus overlap syndrome. J Rheumatol 2018;45(10):1406–1410. doi:https://doi.org/10.3899/jrheum.170953
15. Centers for Disease Control and Prevention. How to protect yourself and others. March 10, 2025. Accessed August 6, 2025. https://www.cdc.gov/covid/prevention/index.html

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1195–1200
DOI 10.1002/acr.80022
Published 2026. This article is a U.S. Government work and is in the public domain in the USA. *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

BRIEF REPORT

Rheumatologic Manifestations of Patients With Type B Insulin Resistance

S. Amara Ogbonnaya, MD,[1] Sandra G. Williams, MD, PhD,[1] Raphael A. Kirou, BS,[1] Marissa Lightbourne, MD, MPH,[2] and Rebecca J. Brown, MD, MHSc[2]

Objective. The objectives of this study were to identify laboratory and clinical features associated with type B insulin resistance (TBIR), a rare condition caused by autoantibodies that inhibit the insulin receptor, most frequently occurring in the setting of systemic lupus erythematosus (SLE), and to increase awareness of this rare, life-threatening condition.

Methods. Thirty-eight patients with TBIR who were seen at the National Institutes of Health between 1976 and 2024 were included. Retrospective chart review was performed to assign primary rheumatologic diagnoses, characterize endocrinologic laboratory measurements, and identify clinical and laboratory manifestations of SLE, including hypocomplementemia, cytopenias, and autoantibody seropositivity.

Results. SLE was the most frequent underlying diagnosis (81.5%); one patient each had Sjögren disease and primary biliary cholangitis. Patients were predominantly female (89.5%) and Black/African American (84.2%). The median Systemic Lupus Erythematosus Disease Activity Index 2000 score was 14 (interquartile range 7). High or very high U1 RNP autoantibodies were seen in >50% of patients. TBIR was associated with a high prevalence of acute neuropathies of the seventh and/or eighth cranial nerve (18.4%), angioedema (10.5%), and uveitis (5%).

Conclusion. TBIR can be a rare complication of SLE, is associated with the presence of high-titer U1 RNP autoantibodies, and may co-occur with other rare SLE manifestations.

INTRODUCTION

Type B insulin resistance (TBIR) is a rare life-threatening autoimmune disorder resulting from production of polyclonal antibodies against the insulin receptor. These autoantibodies block the receptor,[1] leading to extreme insulin resistance that is more severe compared to other diabetes subgroups, often requiring thousands of units of insulin per day.[2] Patients typically present with severe hyperglycemia and weight loss from a highly catabolic state. Other common presenting features include extensive acanthosis nigricans, sometimes involving the periocular and perioral regions, and hyperandrogenism with polycystic ovaries in women of reproductive age.[3]

A case series and a systematic review of 119 TBIR cases identified co-occurring autoimmune disorders in >100 patients with TBIR[3,4]; systemic lupus erythematosus (SLE) was the most commonly associated condition. SLE is a systemic autoimmune disease with variable clinical manifestations that can affect any organ system. It is associated with a diverse array of autoantibodies.[5,6] Mixed connective tissue disease (MCTD), type 1 diabetes mellitus, Hashimoto's thyroiditis, Sjögren disease, and rheumatoid arthritis can also co-occur with TBIR, as can malignancy.[4]

Despite its strong association with rheumatic diseases, rheumatologic manifestations in TBIR have not been well characterized. Thus, the goal of this manuscript is to describe the

This research was supported by the Intramural Research Program of the NIH. The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the United States Department of Health and Human Services.

[1]National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH, Bethesda, Maryland; [2]National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, Maryland.

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.80022).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80022.

Address correspondence via email to Rebecca J. Brown, MD, MHSc, at brownrebecca@niddk.nih.gov.

Submitted for publication August 11, 2025; accepted in revised form January 22, 2026.

SIGNIFICANCE & INNOVATIONS

- This study increases awareness of type B insulin resistance (TBIR), a rare autoimmune condition most frequently occurring in patients with systemic lupus erythematosus (SLE).
- Patients with TBIR and SLE have high disease activity as measured by the Systemic Lupus Erythematosus Disease Activity Index 2000 score, and TBIR can be a presenting manifestation of SLE.
- There is an association of TBIR with acute cranial neuropathies of the seventh and eighth nerve.
- There was a high prevalence of high-titer U1 RNP autoantibodies in patients with TBIR.

autoimmune clinical and laboratory features of patients with TBIR and to increase awareness of the diagnosis.

PATIENTS AND METHODS

Patients and classification criteria. Patients were enrolled in studies approved by the institutional review board of the National Institutes of Health (NIH). Written informed consent or assent was obtained from all patients and/or their parents or guardians.

All patients with TBIR seen at the NIH between 1976 and 2024 were assessed for eligibility, and they were included if they had anti-insulin receptor antibodies measured by a semiquantitative or quantitative assay.[7,8] All but one patient had active TBIR during their evaluation at the NIH. Of 44 patients identified, 4 were excluded due to inaccessible laboratory records (year of diagnosis before electronic medical records and/or illegibility of written records). Two were excluded due to a neoplastic process driving TBIR (plasma cell dyscrasia and lymphoma). Some patients have been reported previously but without a primary focus on rheumatologic manifestations.[2,3,9]

Data collection. Data were collected by retrospective chart review. Race and ethnicity were self-reported. Clinical manifestations of autoimmune diseases were recorded, including cutaneous lupus, arthritis or arthralgias, serositis, lupus nephritis, oral and nasal ulcers, neurologic signs or symptoms, hemolytic anemia, lymphadenopathy, angioedema, and uveitis. Daily insulin requirements were recorded.

Anti-insulin receptor and SLE autoantibody titers (antinuclear autoantibodies [ANAs], double-stranded DNA [dsDNA] autoantibodies, extractable nuclear antigen autoantibodies) were recorded when available. To compare across different testing modalities used during the long study period, autoantibody titers were categorized as ordinal values: titers were categorized as low, moderate, high, or very high.

Other pertinent laboratory values were recorded from the visit with the lowest C3 complement level (or lowest total complement if C3 was not available) as a proxy for peak autoimmune disease activity. Laboratory data closest in time to the complement measurement, but no more than one month before or after, were recorded. For one patient in whom complement levels were not measured, the earliest measurement for each value was recorded (all within a one-month time frame). Laboratory data included measures of insulin resistance (glycosylated hemoglobin [HbA_{1c}], fasting insulin, glucose, and triglyceride levels), measures of inflammation (complement, leukocyte counts, platelet counts, hemoglobin, erythrocyte sedimentation rate, C-reactive protein, albumin, total protein), measures of kidney function (creatinine, blood urea nitrogen, urine protein/creatinine ratio, urine protein excretion), and measures of hormones (growth hormone, insulin-like growth factor 1, parathyroid hormone, osteocalcin, dehydroepiandrosterone sulfate, androstenedione, and testosterone).

Mean and SD were calculated for each laboratory value. The total number of patients included in each analysis varied due to missing data and is noted in each table. Nonquantitative measurements (eg, above or below a measurement threshold) were not included in the mean and SD.

Patients were classified as having SLE if they met 2012 Systemic Lupus International Collaborative Clinics (SLICC) classification criteria before or at the time of presentation with TBIR. We did not award points for alopecia because hyperandrogenism observed in women with TBIR could also cause alopecia. Neurologic findings, such as headaches that could be attributed to other causes, were not awarded points. Psychosis and behavioral changes were not awarded points if they could be attributed to glucocorticoid use, but seizures and cranial neuropathies were included. Patients were classified as having undifferentiated connective tissue disease if they did not meet SLICC classification criteria but there was clinical suspicion for an underlying autoimmune disease and they had autoantibodies.

The Systemic Lupus Erythematosus Disease Activity Index 2000 (SLEDAI-2K), excluding alopecia and neurologic findings that could be attributed to glucocorticoids, was calculated retrospectively to assess disease activity in patients with SLE.

RESULTS

Patient characteristics. Characteristics of the 38 patients are presented in Table 1. Most patients were female (89.5%) and Black/African American (84.2%). Mean age at diagnosis of TBIR was 42 years (range 10.5–64.3 years). SLE was the most frequent underlying diagnosis (81.5%). One patient had primary Sjögren disease, and one had immune-mediated thrombocytopenia and co-existing primary biliary cholangitis confirmed with liver biopsy.

Clinical and laboratory data. At the time of lowest C3 level, 28 patients (76%) were hyperglycemic, 5 (14%) were

Table 1. Patient characteristics and clinical manifestations of the 38 patients with TBIR*

Patient characteristics and clinical manifestations	Values
Age at presentation, mean (range), y	42.4 (10.5–64.3)
Race, n (%)	
Black/African American	32 (84.2)
Hispanic or Latino	4 (10.5)
Not reported	1 (2.6)
White	1 (2.6)
Sex, n (%)	
Female	34 (89.5)
Male	4 (10.5)
Associated autoimmune disorder, n (%)	
SLE	31 (81.5)
UCTD	5 (13.1)
Primary biliary cholangitis	1 (2.6)
Sjögren disease	1 (2.6)
Immune-mediated thrombocytopenia	1 (2.6)[a]
Clinical manifestations, n (%)	
Acute/chronic cutaneous lupus	10 (26.3)
Arthritis/arthralgias	20 (52.6)
Serositis	9 (23.7)
Renal	17 (44.7)
Biopsy-proven lupus nephritis	8 (21.1)
Protein/creatinine ratio > 0.5 mg/mg	9 (23.7)
Neurologic	15 (39.5)
Cranial nerve palsy	7 (18.4)
Psychosis/behavior change	4 (10.5)
Seizures	2 (5.3)
Catatonia	1 (2.6)
Hemolytic anemia	4 (10.5)
Lymphadenopathy	11 (28.9)
Angioedema	4 (10.5)
Uveitis	2 (5.3)

* SLE, systemic lupus erythematosus; TBIR, type B insulin resistance; UCTD, undifferentiated connective tissue disease.
[a] The same patient with primary biliary cholangitis had immune-mediated thrombocytopenia.

Table 2. Pertinent laboratory values of patients*

Measurement	Mean (SD)	n/N (%)
Glucose, mg/dL	174 (109)	
Low (hypoglycemia): <70		4/38 (11)
High (hyperglycemia): >100		29/38 (76)
HbA_{1c}, %[a]	10.6 (3.0)	
Prediabetic: 5.7–6.4		0/33 (0)
Diabetic: ≥6.5		32/33 (97)
Triglycerides, mg/dL	72 (61)	
High (hypertriglyceridemia): ≥150		3/35 (9)
Insulin, μU/mL	1,455 (1,775)	
High: >20		30/31 (97)
C3 complement, mg/dL	63 (20)	
Low[b]		27/35 (77)
C4 complement, mg/dL	16 (8)	
Low[b]		14/35 (40)
Total complement (CH50), U/mL	93 (48)	
Low[b]		6/33 (18)
Leukocyte count, 1,000 cells/μL	3.7 (1.9)	
Low (leukopenia): <4		25/38 (66)
Lymphocyte count, 1,000 cells/μL	1.1 (0.5)	
Severe lymphopenia: <0.5		3/34 (9)
Moderate lymphopenia: 0.5–1		14/34 (41)
Mild lymphopenia: 1–1.5		9/34 (26)
Platelet count, 1,000 cells/μL	213 (89)	
Severe thrombocytopenia: <50		0/38 (0)
Moderate thrombocytopenia: 50–100		2/38 (5)
Mild thrombocytopenia: 100–150		4/38 (11)
Thrombocytosis: >450		1/38 (3)
Urine protein/creatinine ratio, mg/mg	0.75 (1.09)	
High: ≥0.2		15/21 (71)
Urine protein excretion, g/24 h	1 (1.1)	
Total proteinuria: ≥0.5		14/30 (47)
Nephrotic-range proteinuria: >3.5		3/30 (10)

* The subrow headings represent abnormal reference ranges and are mainly derived from the American Board of Internal Medicine (updated January 2025). HbA_{1c}, glycosylated hemoglobin; CH50, total hemolytic complement.
[a] One patient had an HbA_{1c} measurement of 4.5%, but the value was excluded due to hemolytic anemia.
[b] Due to substantial differences in reference ranges between complement measurements, low values were assigned for each measurement separately using the corresponding reference range listed in the record.

euglycemic, and 4 (10.5%) were hypoglycemic. Mean fasting glucose was 176 mg/dL. Mean HbA_{1c} was 10.6%, with 32 of 33 patients (97%) in the diabetic range (Table 2). As expected, serum insulin levels were elevated (>20 μU/mL) in 30 of 31 patients (97%), with a mean of 1,455 μU/mL and a median of 786 μU/mL (Table 2). Only 3 of 35 patients (9%) exhibited hypertriglyceridemia (≥150 mg/dL), with a mean triglyceride level of 72 mg/dL (Supplementary Table 1, Table 2). The median exogenous insulin dose was 2,400 units/day (Supplementary Table 2).

A majority exhibited leukopenia, lymphopenia, and anemia, with 58% of patients having bicytopenia or pancytopenia (Table 2). A minority of patients had elevated creatinine or azotemia, but 14 of 30 (47%) had proteinuria; 3 (10%) were in the nephrotic range (Table 2). Patients also exhibited abnormalities in hormone levels, most notably high total testosterone in women (Supplementary Table 3).

Regarding SLE autoantibodies, ANA testing was positive in >90% of patients (Table 3). Positive U1 RNP autoantibodies were seen in >70%, with the majority having high or very high titers (Table 3).

Of the 31 patients classified as having SLE, 84% had SLEDAI-2K scores of ≥10, indicating high disease activity (Supplementary Figure 1). Among all patients, arthritis or inflammatory arthralgias was common (20 patients; Table 1) and most frequently involved the hands and wrists. Evidence of renal involvement in lupus was found in 17 patients, either by presence of biopsy-proven lupus nephritis or by presence of a protein/creatinine ratio of >0.5 mg/mg (Table 1). Of the 11 patients reported to have diffuse lymphadenopathy, 8 had biopsies; all available pathology reports described findings of reactive lymphadenopathy. Eleven patients had bone marrow biopsies for cytopenias before or at the time of presentation with TBIR, which were benign. Two patients had a history of uveitis, and four experienced angioedema during presentation for TBIR (Table 1).

Table 3. Autoantibodies in patients with TBIR*

Autoantibody	Patients, n (% of autoantibody group)
ANA	38
positive	35 (92.1)
very high titer	20 (52.6)
high titer	5 (13.2)
moderate titer	4 (10.5)
low titer	6 (15.8)
negative	3 (7.9)
U1 RNP	32
positive	23 (71.8)
very high titer	13 (40.6)
high titer	4 (12.5)
moderate titer	0 (0)
low titer	1 (3.1)
unknown titer	5 (15.6)
negative	9 (28.1)
dsDNA	34
positive	11 (32.4)
negative	23 (67.6)
Smith	11
positive	5 (45.5)
negative	6 (54.5)
SSA	22
positive	12 (52.1)
negative	10 (47.8)
SSB	22
positive	4 (21.7)
negative	18 (78.2)
Jo1	14
positive	0 (0)
negative	14 (100)

* The majority of patients had very high-titer levels U1 RNP autoantibody. ANA titers were categorized as low (<1:320 or 1 to <4 EU), moderate (>1:320 to 1:640 or 4 to <8 EU), high (>1:640 to <1:1,280 or 8–12), or very high (1:1,280 or >12). Anti–U1 RNP titers were categorized as negative (<19 EU or <1 U), low (20–80 EU or 1.1 to <3 U), moderate (80–140 EU or 3–6 U), high (140–200 EU or 6–8 U), or very high (>200 EU or >8 U). ANA, antinuclear autoantibody; dsDNA, double-stranded DNA; EU, enzyme units; TBIR, type B insulin resistance.

Neurologic symptoms occurred in 15 patients (Table 1). There was a high rate of acute cranial neuropathies of the seventh and/or eighth nerve, including unilateral or bilateral facial droop, and stapedius dysfunction and/or sensorineural hearing loss confirmed on audiometric evaluation, which occurred in seven patients (Table 1). Four patients had behavioral changes or psychosis. In two patients, this occurred before initiation of glucocorticoids, but in the remaining two, it was unclear if symptoms started before glucocorticoid initiation. Systemic autoimmune disease diagnoses occurred before diagnosis of TBIR in approximately 35% and following the onset of TBIR in 65%. The median time between onset of TBIR and another autoimmune diagnosis was 2.7 months (interquartile range 26.5 months; Supplementary Figure 2).

DISCUSSION

TBIR is characterized by polyclonal antibodies to the insulin receptor and can occur in association with SLE. Testing for insulin receptor autoantibodies is not widely available but can be obtained on a research basis.[7,8] Anti-insulin receptor autoantibodies behave as partial agonists; at high titers, the antibodies antagonize the insulin receptor, resulting in hyperglycemia and severe insulin resistance in the absence of other inciting factors such as glucocorticoids.[1,4] At low titers, these antibodies stimulate the insulin receptor, resulting in hypoglycemia, and may mimic an insulinoma.[2] In our cohort, almost all patients had very high insulin requirements, with the exception of patients who presented with recurrent, life-threatening hypoglycemia. Patients with TBIR may present to rheumatologists or endocrinologists, and effective management requires close collaboration between these subspecialties. The presence of TBIR may influence the choice of SLE-directed therapy, as combination therapy with anti-CD20 agents plus additional immunomodulatory agents is more effective than monotherapy.[2]

Patients with TBIR have very high fasting serum insulin, elevated HbA_{1c}, elevated adiponectin, low insulin-like growth factor 1, negative insulin autoantibodies, and positive insulin receptor autoantibodies. As opposed to type 2 diabetes, in which serum triglycerides are high, triglycerides are low in TBIR, consistent with the findings in our cohort. These laboratory tests may be useful in evaluating patients with suspected TBIR while awaiting anti-insulin receptor autoantibody levels.

Autoantibody testing revealed positive ANA results in all but one patient with TBIR and SLE. We suspect this could be due to the prozone or "hook effect," in which very high antibody titers cause a falsely negative result. Very high anti–U1 RNP titers were associated with TBIR, whereas anti-dsDNA autoantibody results were negative in two-thirds, and anti-Smith autoantibody results were positive in 50%. U1 RNP autoantibody positivity has been reported in approximately 20% of patients newly diagnosed with lupus in some studies.[5] In our TBIR cohort, four patients who met SLICC classification criteria for SLE had a clinical diagnosis of MCTD in the medical record. However, we were not able to accurately elucidate the frequency of Raynaud phenomenon, nailfold capillary changes, or muscle or pulmonary manifestations from available records. We did not identify a signal for clinical features of MCTD in patients with U1 RNP–positive TBIR. Longitudinal assessment of patients with U1 RNP–positive TBIR is needed to understand whether they develop features of overlap syndromes. Given the association between U1 RNP antibodies and severe Raynaud phenomenon, pulmonary hypertension, and interstitial lung disease, it would be prudent to perform longitudinal screening.

Our results show that in most cases, onset of TBIR occurs in close proximity with onset of SLE, as diagnoses of SLE and TBIR usually occurred within a few months of each other (Supplementary Figure 2). However, the diagnoses may be challenging, as some patients come to medical attention for insulin resistance and its complications with delayed recognition of systemic manifestations of SLE. Other patients may come to

attention because of classic SLE symptoms and are found to have diabetes. In these patients, limited awareness of TBIR and difficulties obtaining confirmatory diagnostic testing may delay diagnosis.

Patients with SLE and TBIR have high SLE disease activity. Over 80% had SLEDAI-2K scores ≥10, which is likely underestimated given incomplete clinical data. Common clinical manifestations of SLE in TBIR included lymphadenopathy, cytopenias, arthritis, renal disease, and acute and chronic cutaneous lupus. These findings are common in the general SLE population, with lymphadenopathy occurring frequently during SLE flares, lymphopenia occurring in >90%, renal involvement in up to 50%, musculoskeletal involvement in up to 90%, and skin involvement and/or photosensitivity in 90%.[6] These clinical features occurred at similar frequencies in the TBIR cohort with the exception of cutaneous manifestations, which were underrepresented.

Interestingly, there was an unusually high prevalence of cranial neuropathies (Table 1). Cranial neuropathies are uncommon manifestations of SLE, with prior studies reporting occurrence in <2% of patients with SLE.[10] In our TBIR cohort, cranial neuropathies were present in 18.4%. There is a known association between Bell's palsy and diabetes. The mechanism is unknown, but it is suggested that Bell's palsy is the result of nerve ischemia.[11] In association with TBIR, it is unclear if the etiology of cranial neuropathies is autoimmune, a microvascular effect of diabetes, or a combination of both processes. We suspect an immune-mediated process may contribute as there are associations with specific autoantibodies and cranial neuropathies in lupus.[12] Although there was a higher frequency of psychosis and behavioral changes, we could not exclude the contribution of glucocorticoid use in some patients. Our findings suggest that there should be a high index of suspicion for neurologic manifestations in patients presenting with TBIR.

Anterior uveitis is very rarely associated with SLE, with reported prevalence of 0.6% to 0.8% in patients with adult and juvenile onset SLE, respectively.[13] In our cohort, 2 of 38 (5.3%) patients with TBIR and autoimmune disease had a history of anterior uveitis at the time of, or before, TBIR diagnosis. Angioedema is an infrequent but well-established manifestation of SLE and is associated with an acquired C1q esterase inhibitor deficiency causing unmediated bradykinin signaling. Angioedema occurred in 13.2% of patients with TBIR; C1q esterase activity was not measured.

Management of TBIR involves treating the underlying autoimmune disease and high-dose insulin to reverse the hypercatabolic state.[2] Extremely high doses of insulin may be required, sometimes via continuous infusion or concentrated subcutaneous insulin, and normalization of blood glucose is frequently impossible early in treatment. Optimal immunomodulatory therapy is not known; however, a combination of medications up front seems to be more efficacious than monotherapy. High rates of remission were observed with cyclophosphamide, glucocorticoids, and anti-CD20 monoclonal antibody combination therapy.[2] As in the general population, high-dose glucocorticoids may acutely worsen insulin resistance in TBIR. However, high-dose glucocorticoids are associated with decreased immunoglobulins and also affect B cell proliferation and differentiation to plasma cells,[14] ultimately contributing to resolution of insulin resistance and diabetes. Both autoimmune features and manifestations of insulin resistance, including hyperglycemia, acanthosis nigricans, and hyperandrogenism, improve with treatment. TBIR recurs in a minority of patients after treatment.[3] Given the promising initial results of chimeric antigen receptor T cell therapy in autoimmune diseases,[15] this may be a consideration in patients with refractory disease or life-threatening hypoglycemia.

In conclusion, TBIR is a rare condition that typically occurs concomitantly with SLE. Although many clinical and laboratory manifestations of connective tissue disease in patients with TBIR were typical for patients with SLE, the presence of high-titer positive U1 RNP autoantibodies, higher incidence of cranial neuropathies, anterior uveitis, and angioedema may be distinguishing features of SLE associated with TBIR. A multidisciplinary team approach including endocrinology, rheumatology, and the assistance of a skilled pharmacist for appropriately administering insulin is extremely helpful in managing these patients.

This study was limited by the retrospective chart review design with consequent missing data, as well as lack of a control group with SLE without TBIR and difficulty comparing autoantibodies over time due to laboratory methodologic changes. Longitudinal rheumatologic evaluation of patients with TBIR would be of interest, particularly regarding development of pulmonary hypertension given the strong association with U1 RNP autoantibodies. Additional immunoprofiling for autoantibodies associated with cranial neuropathies to identify associated autoantibodies and genetic evaluation of patients with TBIR may be of further interest.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Brown 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. Flier JS, Kahn CR, Roth J, et al. Antibodies that impair insulin receptor binding in an unusual diabetic syndrome with severe insulin resistance. Science 1975;190(4209):63–65. doi:https://doi.org/10.1126/science.170678
2. Klubo-Gwiezdzinska J, Lange M, Cochran E, et al. Combined immunosuppressive therapy induces remission in patients with severe type B insulin resistance: a prospective cohort study. Diabetes Care 2018; 41(11):2353–2360. doi:https://doi.org/10.2337/dc18-0884

3. Arioglu E, Andewelt A, Diabo C, et al. Clinical course of the syndrome of autoantibodies to the insulin receptor (type B insulin resistance): a 28-year perspective. Medicine (Baltimore) 2002;81(2):87–100. doi: https://doi.org/10.1097/00005792-200203000-00001
4. Martins LM, Fernandes VO, Carvalho MMD, et al. Type B insulin resistance syndrome: a systematic review. Arch Endocrinol Metab 2020; 64(4):337–348. doi:https://doi.org/10.20945/2359-3997000000257
5. Li WX, Pan HF, Li LH, et al. Prevalence and clinical significance of 15 autoantibodies in patients with new-onset systemic lupus erythematosus. Ir J Med Sci 2010;179(4):623–627. doi:https://doi.org/10.1007/s11845-010-0543-x
6. Dall'era M. Clinical features of systemic lupus erythematosus. In: Firestein GS, McInnes IB, Koretzky G, et al, eds. Firestein & Kelley's Textbook of Rheumatology. 12th ed. Elsevier;2025:1404–1428.e5.
7. Minich WB, Abel BS, Schwiebert C, et al. A novel in vitro assay correlates insulin receptor autoantibodies to fasting insulin in type B insulin resistance. J Clin Endocrinol Metab 2023;108(9):2324–2329. doi: https://doi.org/10.1210/clinem/dgad125
8. Geng L, Wong CL, Liao B, et al. Development of a novel diagnostic assay for insulin receptor autoantibodies based on a patient with autoimmune hypoglycaemia. Front Endocrinol (Lausanne) 2022;13: 1029297. doi:https://doi.org/10.3389/fendo.2022.1029297
9. Malek R, Chong AY, Lupsa BC, et al. Treatment of type B insulin resistance: a novel approach to reduce insulin receptor autoantibodies. J Clin Endocrinol Metab 2010;95(8):3641–3647. doi:https://doi.org/10.1210/jc.2010-0167
10. Muscal E, Brey RL. Neurologic manifestations of systemic lupus erythematosus in children and adults. Neurol Clin 2010;28(1):61–73. doi:https://doi.org/10.1016/j.ncl.2009.09.004
11. Adour K, Wingerd J, Doty HE. Prevalence of concurrent diabetes mellitus and idiopathic facial paralysis (Bell's palsy). Diabetes 1975;24(5): 449–451. doi:https://doi.org/10.2337/diab.24.5.449
12. Krustev E, Hanly JG, Chin R, et al. Anti-KIF20B autoantibodies are associated with cranial neuropathy in systemic lupus erythematosus. Lupus Sci Med 2024;11(1):e001139. doi:https://doi.org/10.1136/lupus-2023-001139
13. Tani C, Elefante E, Arnaud L, et al. Rare clinical manifestations in systemic lupus erythematosus: a review on frequency and clinical presentation. Clin Exp Rheumatol 2022;40 Suppl 134(5):93–102. doi: https://doi.org/10.55563/clinexprheumatol/jrz47c
14. Strehl C, Ehlers L, Gaber T, et al. Glucocorticoids-all-rounders tackling the versatile players of the immune system. Front Immunol 2019; 10:1744. doi:https://doi.org/10.3389/fimmu.2019.01744
15. Müller F, Taubmann J, Bucci L, et al. CD19 CAR T-cell therapy in autoimmune disease - a case series with follow-up. N Engl J Med 2024; 390(8):687–700. doi:https://doi.org/10.1056/NEJMoa2308917

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1201–1208
DOI 10.1002/acr.80001
© 2026 American College of Rheumatology

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

Conflicts of Interest Reporting in Trials and Guidelines Addressing Glucocorticoid Injections for Knee Osteoarthritis

Craig Vecchiarelli, PT, DPT,[1] Lee Newman, PT, MPT,[1] Shannon T. Boyd, PT, DPT,[1] Jodi L. Young, PT, DPT, PhD,[1] and Daniel I. Rhon, PT, DPT, DSc, PhD[2]

Objective. This study aimed to characterize conflict of interest disclosure practices in trials and guidelines recommending glucocorticoid injections for knee osteoarthritis.

Methods. Three databases (CINAHL, Ovid MEDLINE ALL, and Embase) were queried for randomized controlled trials from database inception to April 2025 that assessed glucocorticoid injection treatment effect for knee osteoarthritis. Clinical practice guidelines were retrieved from a recent systematic review. Study details, authors, affiliations, and conflict of interest disclosures were extracted. Transparency and appropriateness with addressing disclosures were assessed for every article, and disclosures were examined and compared with three national public conflict of interest disclosure databases. All conflicts were categorized and proportions calculated.

Results. Seventy-five trials and 14 guidelines were included. Twenty-nine percent of trials (n = 22) and 14.3% of guidelines (n = 2) had no conflict of interest statement. Ten trials (13.3%) and six guidelines (42.9%) reported a conflict of interest for at least one author. Eleven trials (14.7%) and six guidelines (42.9%) had discrepancies between disclosures in articles and reports in public databases. Forty-three trial authors (34.1%) and 19 (9.4%) guideline authors had discrepancies between disclosures in the articles and public databases.

Conclusion. Conflict of interest reporting practices in trials and guidelines assessing effectiveness of glucocorticoid injections for knee osteoarthritis are poor, with a lack of transparency. Quality and thoroughness with reporting conflicts of interest is necessary to best understand industry influence on treatment recommendations.

INTRODUCTION

Intraarticular glucocorticoid injections are one of the most common nonsurgical treatments for knee osteoarthritis, recommended in multiple clinical practice guidelines.[1–6] Between 43% and 50% of patients who seek care for knee osteoarthritis will receive a glucocorticoid injection,[7,8] and it is not uncommon to receive more than one.[7,9,10] Glucocorticoids have demonstrated short-term efficacy[11] and are recommended in several guidelines for various lengths of symptomatic relief.[3,4] However, some have called into question the benefits of glucocorticoid injections, specifically the lack of long-term therapeutic effect weighed against potential risks.[12,13] Observational studies have expressed concern that intraarticular glucocorticoids may contribute to an increased risk for osteoarthritis progression and eventual total knee arthroplasty.[9,14] Several randomized trials found no benefit for glucocorticoid injection compared to physical therapy or exercise alone.[15–17] Another randomized trial demonstrated a significant loss of total knee cartilage volume with glucocorticoids compared to those receiving saline injections.[13] The most recent Cochrane review regarding glucocorticoid injections concluded only low-quality evidence with unclear clinically meaningful benefit,[18] and recent trials show other less invasive treatments to be safer and more effective.[15,19] In a systematic review of 25 clinical practice guidelines, four high-quality guidelines included glucocorticoid injection recommendations for knee osteoarthritis.[20] One of these guidelines provided a strong unconditional recommendation for

The view(s) expressed herein are those of the author(s) and do not necessarily reflect the official policy or position of the Defense Health Agency, Brooke Army Medical Center, the Department of Defense, the Uniformed Services University, nor any agencies under the US Government.

[1]Bellin College, Green Bay, Wisconsin; [2]Uniformed Services University, Bethesda, Maryland.

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.80001).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80001.

Address correspondence via email to Shannon T. Boyd, PT, DPT, at shannon.boyd@bellincollege.edu.

Submitted for publication July 5, 2025; accepted in revised form December 17, 2025.

SIGNIFICANCE & INNOVATIONS

- There is a lack of transparency with conflict of interest (COI) reporting in trials and guidelines for use of glucocorticoid injections for knee osteoarthritis.
- Better and more thorough COI reporting practices could improve assessment of industry influence on trial outcomes and guideline recommendations.

glucocorticoid injections,[3] and the other three provided conditional recommendations.[4,5,21]

Conflicts of interest (COIs) occur when secondary interests such as financial gain and career advancement unduly influence the execution, interpretation, or conclusion of a study.[22] COIs may create bias that influences recommendations, making proper disclosure critical to transparency.[23,24] For example, conflicts were found for authors of clinical practice guidelines for opioid-prescribing practices, possibly driving the overuse of opioid prescriptions.[25,26] Thirteen guidelines published at the height of the opioid epidemic had pervasive COIs with the pharmaceutical industry.[26,27] In general, the quality of COI reporting is poor in orthopedic research.[23,28–30] Data from 2014 show that there were 5,274 published trials investigating glucocorticoid injections with industry funding compared to only 873 trials (16.6% of all trials) published with government funding,[31,32] and a 2017 Cochrane review demonstrated that industry-sponsored research tends to demonstrate more favorable treatment efficacy compared to non–industry-sponsored research.[33]

To date, an assessment of potential COIs driving the recommendations for the use of glucocorticoid injections for knee osteoarthritis has not been properly explored. Therefore, the purpose of this study was to explore COI reporting practices in trials and guidelines assessing the effect of glucocorticoid injections for knee osteoarthritis and determine the accuracy of the reports by comparing disclosures of COIs in study reports to those in publicly available databases.

METHODS

Our approach required (1) a systematic process to identify all relevant literature, (2) an assessment of quality and transparency of COI disclosure in relevant trials and guidelines, and (3) assessment of the relationship between the presence of COIs and recommendations for use of glucocorticoid injections in patients with knee osteoarthritis. Therefore, the literature was systematically searched to identify the scope to which COI disclosures were present and reported in knee osteoarthritis trials.

Scoping review protocol and registration. A scoping review methodology was employed to ensure thorough identification of available literature and followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews checklist.[34] The protocol for this study was registered a priori in the Open Science Framework registry (https://osf.io/sd96h/?view_only=59857156d28640e4bf196dc59b6f864b).

Search strategy. A comprehensive literature search was conducted initially through May 2024 and then updated through April 2025, using three primary databases: CINAHL, Ovid MEDLINE ALL, and Embase. Refer to Supplementary S2 for Ovid MEDLINE ALL search string. The search aimed to identify randomized controlled trials comparing glucocorticoid injections for knee osteoarthritis to placebo or comparator treatments from database inception to April 2025. Inclusion criteria consisted of randomized controlled trials using glucocorticoid injections in one of the treatment groups (no cluster randomized trials were identified), trials published in English, and trials assessing outcomes of pain, physical function, or disability outcome measure at baseline and at least one follow-up time point. Clinical practice guidelines from a recent systematic review of practice guidelines for hip and knee osteoarthritis were also included.[20] Only guidelines that addressed glucocorticoid injection use for knee osteoarthritis were included in our assessment.

Study selection. Screening and data extraction were performed using Covidence, a web-based collaboration program allowing for storage, management, and data extraction of studies for systematic and scoping reviews.[35] Two independent reviewers (CV and LN) screened titles and abstracts for inclusion, and any differences in judgment were resolved by a third reviewer (STB). Full-text trials were retrieved and assessed for eligibility by two reviewers (CV and STB), with any disagreements resolved by a third reviewer (LN).

Data extraction. The following information was extracted from each trial and guideline: title, author names, year of publication, presence of COI statement, COI disclosures for every author, use of standardized process or form for reporting COIs such as the International Committee of Medical Journal Editors (ICMJE),[36,37] funding sources, and treatment effect (favoring glucocorticoid injections or comparator or null effect) for each trial or recommendation for glucocorticoid injection use from each guideline. In some cases, the actual COI disclosure does not appear in the article, but rather, the reader is directed to a link on the journal webpage in which disclosures related to a given article are stored. These sources were also searched to ensure thoroughness.

Twenty trials were independently reviewed by three reviewers (CV, STB, and LN) to ensure reliability with data extraction. If agreement was <80%, another 10 trials were assessed by each reviewer until percentage agreement reached 80%. Any unresolved conflicts were resolved by consensus. The remaining

trials were then divided among three reviewers (CV, STB, and LN) for data extraction. Reviewers were not blinded to study results or author names.

Open COI disclosure databases were searched for records of authors' COIs between May and July 2024 for the initial search and again in April 2025 for the updated search. Data from three databases were collected: the American Academy of Orthopaedic Surgeons Disclosure database (https://disclosure.aaos.org/search), the Centers for Medicare and Medicaid Services Open Payments database (https://openpaymentsdata.cms.gov/), and the ProPublica Dollars for Profs database (https://projects.propublica.org/dollars-for-profs/). For each author, the year, dollar amount, and reported COI disclosures were recorded.

Data synthesis and analysis. A descriptive synthesis was used to pool data focused on COI disclosures and trial or guideline recommendations for use of glucocorticoid injections. The results of each trial were also assessed to categorize each trial based on whether a positive treatment effect favoring glucocorticoid injections was identified (YES/NO). Similarly, each clinical practice guideline was characterized based on its recommendation for the use of glucocorticoid injections (YES/NO). Self-reported COIs were identified from each publication when reported and compared to COIs collected from the three databases. Summary statistics were calculated for the number and type of COIs categorized as stock ownership, company ownership, patents, or royalties; paid speaker, presenter, or consultant; research support; employment or other financial interests; committee or board member; travel or lodging; food or beverage; other professional activity; education; entertainment; gift; and debt forgiveness. From these categories, we specifically assessed the following COIs: company ownership, stock ownership, royalties, or patents; professional leadership roles including committee or board member positions; paid speaker, presenter, or consultant; employment or other financial interests; and research support. COIs were categorized as being present within three years before publication or not. The three-year time frame was chosen based on recommendations for COI disclosure from the ICMJE.[36] COI categories found in the databases were compared to publication disclosures. If there were publication disclosures and public databases that did not align, it was considered a discrepancy.

RESULTS

The search resulted in 1,146 trials after removal of duplicates and nonrandomized trials. After title, abstract, and full-text screening, 75 trials met the final inclusion criteria (Supplementary Material 2). Reasons for exclusion are shown in Figure 1. There were 475 authors from all trials. Eleven of the 75 trials (14.7%) found glucocorticoid injections to be effective compared to controls.

Fourteen of 25 guidelines (Supplementary Material 3) were selected for extraction after removal of guidelines that were not specific to knee osteoarthritis or that did not include recommendations for glucocorticoid injections. Collectively, there were 352 authors from the guidelines. All 14 guidelines recommend the use of intraarticular glucocorticoid injections to varying degrees.

Trial characteristics. Fifty-three (70.7%) of the 75 trials included a COI statement, and 10 (13.3%) reported a COI for at least one author. Eleven (14.7%) of the 75 trials had authors with COIs within the three years preceding trial publication. All 11 trials had discrepancies between COI statements in the articles and COIs found in the public databases. Refer to Table 1 for glucocorticoid injection trial characteristics and descriptives.

COIs in trials. Forty-three authors (9.1%) had COIs within the three years of trial publication. There were 88 COIs disclosed in the articles and 180 identified COIs found in the public databases. The most notable discrepancies were disclosures observed in the public databases at a rate three times higher than in the articles. For example, there were 20 instances of paid speaker, presenter, or consultant disclosed in the articles yet 63 identified instances of paid speaker, presenter, or consultant in the public databases. Likewise, there were nine instances of stock ownership, company ownership, royalties, or patents disclosed in the publications but 27 similar COIs found in the public databases. The complete breakdown of the relationship of COIs disclosed in publications versus public databases is found in Table 1.

Trial outcomes. Of 11 trials (14.7%) that found glucocorticoid injections to be effective, two were published before public database inception. Four trials were from international authors in international journals that did not have a COI disclosure in the article or one found in the public databases. Another international trial had a COI disclosure in the article but no information found in the databases. Two trials from the United States had COI disclosure statements in the articles but none in the public databases. Lastly, three trials had a COI disclosure statement that differed from the COIs found in the public databases.

Guideline characteristics. Of 14 clinical practice guidelines, 12 (85.7%) included a COI statement in the article, with six (42.9%) disclosing a COI for at least one author. The remaining six guidelines included a statement of "nothing to disclose" or noted an internal management plan for addressing COIs without further details regarding the nature of the conflicts. Ten (71.4%) of the 14 guidelines had at least one author with a COI found in

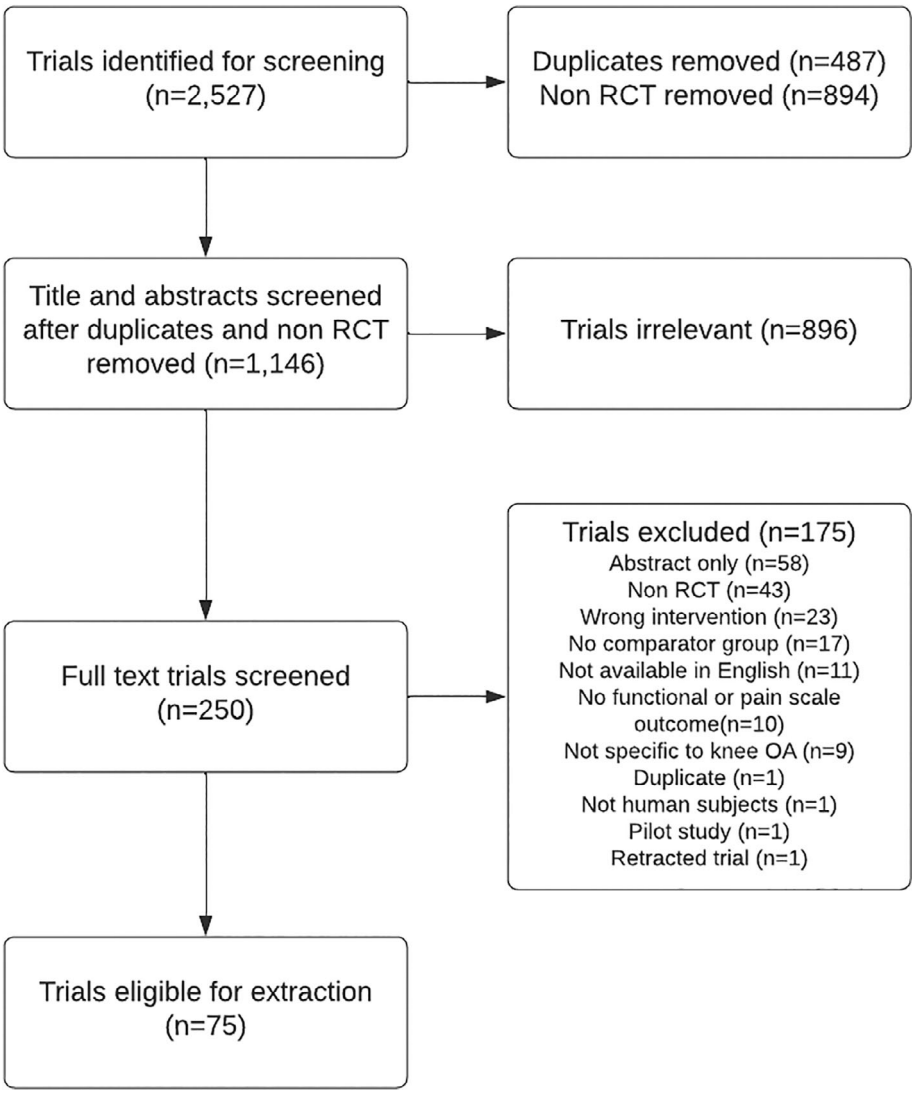

Figure 1. Flow diagram of literature search and selection process. OA, osteoarthritis; RCT, randomized controlled trial.

a public database, with four guidelines (28.6%) having authors who had not disclosed the conflict in the article.

COIs in guidelines. Of the 10 guidelines with COIs reported in a public database, eight (80%) had COIs within the three years before publication. There was a total of 203 authors across these eight guidelines, of whom 19 (9.4%) had discrepancies between COIs found in databases and COIs disclosed in the guidelines. Discrepancies between COIs disclosed in the guidelines and those found in public databases included 15 instances of research support; 10 instances of paid speaker, presenter, or consultant; three instances of stock, company ownership, patents, or royalties; and two instances of committee or board member participation. Refer to Table 2 for glucocorticoid injection guideline descriptives and breakdown of the relationship of COIs disclosed in publications versus public databases.

Guideline recommendations. To varying degrees, all guidelines recommend the use of glucocorticoid injections for conservative care of knee osteoarthritis.

DISCUSSION

Overall, there was substantial variability between COIs disclosed in the articles and those reported in public databases. Twenty-two trials (29.3%) and two clinical practice guidelines (14.3%) lacked a COI statement. Authors in both trials and guidelines had COIs that did not match those found in public databases, or they did not report COIs that were found to be present in public databases.

There were inaccuracies reported in both the trials and guidelines. This variability may be driven by several factors. Physicians may not be aware of reporting standards,[38] and although there is general movement toward a standardized method of

Table 1. Randomized controlled trials

	n (%)
Randomized controlled trials (n = 75)	
Conflicts of interest statement present	53 (70.7)
Conflicts of interest disclosed in publication	10 (13.3)
Conflicts of interest in databases	24 (32)
Conflicts of interest in database disclosed within three years of publication	11 (14.7)
With discrepancy between database and publication	11 (14.7)
Found glucocorticoid injections effective	11 (14.7)
Without conflicts of interest statement present	5 (6.7)
With conflicts of interest statement—nothing to declare	2 (2.7)
With conflicts of interest statement and conflicts to declare	4 (5.3)
All authors from randomized controlled trials (n = 475)	
Conflicts of interest disclosed in publication	65 (13.7)
Conflicts of interest in databases	91 (19.2)
Only authors from randomized controlled trials with conflicts of interest (n = 126)	
Discrepancy between database and publication	43 (34.1)
All categories from randomized controlled trials	
Publication (n = 88)	
Employment or other financial interests	30 (34.1)
Paid speaker, presenter, consultant	20 (22.7)
Research support	15 (17.0)
Committee or board member	14 (15.9)
Stock, ownership, patents, royalties	9 (10.2)
Databases (n = 180)	
Paid speaker, presenter, consultant	63 (35)
Research support	45 (25)
Committee or board member	33 (18.3)
Stock, ownership, patents, royalties	27 (15)
Employment or other financial interest	12 (6.7)
Category discrepancies from trials with conflicts of interest (n = 75)	
Stock, ownership, patents, royalties	14 (18.7)
Paid speaker, presenter, consultant	30 (40)
Research support	23 (30.7)
Employment or other financial interests	3 (4)
Committee or board member	5 (6.7)

Table 2. Clinical practice guidelines*

	n (%)
CPGs (n = 14)	
COI statement present	12 (85.7)
COI disclosed in publication	6 (42.9)
COIs in database	10 (71.4)
COI in database disclosure within three years of publication	8 (57.1)
With discrepancy between database and publication	6 (42.9)
Recommended glucocorticoid injections	14 (100)
All authors from CPGs (n = 352)	
COIs disclosed in publication	61 (17.3)
COIs in databases	55 (15.6)
Only authors from CPGs with COI (n = 203)	
Discrepancy between database and publication	19 (9.4)
All categories from CPGs	
Publication (n = 81)	
Paid speaker, presenter, consultant	41 (50.6)
Committee or board member	20 (24.7)
Research support	13 (16)
Stock, ownership, patents, royalties	6 (7.4)
Employment or other financial interest	1 (1.2)
Not disclosed (n = 95)[a]	95 (100)
Databases (n = 110)	
Paid speaker, presenter, consultant	42 (38.2)
Research support	26 (23.6)
Committee or board member	25 (22.7)
Stock, ownership, patents, royalties	14 (12.7)
Employment or other financial interest	3 (2.7)
Category discrepancies from CPG with COI (n = 30)	
Stock, ownership, patents, royalties	3 (10)
Paid speaker, presenter, consultant	10 (33.3)
Research support	15 (50)
Employment or other financial interests	0 (0)
Committee or board member	2 (6.7)

* COI, conflict of interest; CPG, clinical practice guideline.
[a] Osteoarthritis Research Society International, US Department of Veterans Affairs and US Department of Defense, Malaysia Health Technology Assessment Section—these guidelines referenced a conflicts of interest management plan but did not provide details of conflicts of interest in publication.

reporting COIs, this standardization is not a requirement. Variability in the reporting process from journal to journal may lead to discrepancies (potentially due to human error or perception) of what should be self-reported, especially if an author deems a COI not relevant to the manuscript being submitted.[39] However, this value would still be present in the Open Payments database by the reporting entity (commonly pharmaceutical and medical device companies). Physicians may not be aware of values of certain categories such as food expenses, cost of travel, cost of lodging, or conference expenses when paid for by a reporting entity. This can lead to discrepancies when the majority of conflict reporting in journals is exclusively self-reported, whereas the Open Payments database information is primarily via reporting entities. Lastly, timing of conflict reporting may lead to discrepancies between when a COI was reported in an article and when a COI was uploaded and available in a public database.

Ultimately, it is difficult to determine why discrepancies between article and database COIs exist because there is no universally standardized method of collecting and reporting COIs. The absence of a universally standardized reporting process leaves self-reporting open to interpretation. These factors can contribute to discrepancies identified between articles and databases.

Although systematic variability in reporting leads to discrepancies between articles and public databases, there appeared to be no substantial relationship between COIs and trial outcomes. A low number of trials (n = 11) in this study occurred after database inception, and all of those trials demonstrated discrepancies between COIs in the articles and those reported in the public databases. This may contribute to the percentage of authors identified, with COI discrepancies being less than what has been reported in other orthopedic literature.[40–43] There could be several reasons to explain this. With glucocorticoid injections in the medical literature since the 1960s, there have been many trials

published before public COI disclosure databases were available. Policies to disclose COIs did not become commonplace until the later 1990s or early 2000s.[44] For example, reporting was not made mandatory by *The Journal of the American Medical Association* until 2006.[45] Likewise, the Physician Payment Sunshine Act of 2010 was enacted by legislation in the United States mandating that payments made by pharmaceutical companies and medical device companies to physicians be disclosed in a publicly available database, with data collection commencing in 2013.[46] Additionally, there were several trials and guidelines with international authors. Because COI disclosure databases designed primarily for authors in the United States were used in this review, there may have been discrepancies of international authors that were not identified unless they participated in American Academy of Orthopaedic Surgeons' activities requiring public disclosure in the Academy's COI database. The lack of a centralized reporting system on a global scale limits the ability to investigate COIs thoroughly and implies the need for a solution.

There was a smaller percentage of discrepancies between COIs disclosed in articles and those in public databases for the clinical practice guidelines compared to the trials. This may be due to a greater number of nonphysician researchers participating in clinical practice guidelines or a larger number of international authors involved in clinical practice guideline development, potentially skewing the percentage of discrepancies. International authors and nonphysician authors would not have reported COIs in the Center for Medicare and Medicaid Services Open Payments database because neither group consists of providers in the United States. Likewise, these groups would only have COIs in the American Academy of Orthopaedic Surgeons database if they were participating in American Academy of Orthopaedic Surgeons' activities that required reporting. Therefore, there could potentially be underreporting of self-disclosed conflicts with no other methods for identifying COIs. All clinical practice guidelines recommended the use of glucocorticoid injections for knee osteoarthritis in some capacity.

A small number of trials included in this review reported outcomes favoring the glucocorticoid injections group. However, five of the 11 trials with favorable results for glucocorticoid injections did not have a COI in the article, and three of the seven trials published after public database inception consisted of international authors in international journals, which are not required to report in the databases this study included. No other associations between underdisclosure or nondisclosure of COIs in the articles were found when compared to those in the databases and whether the outcomes favored the effectiveness of glucocorticoid injections for knee osteoarthritis.

COI statements must accompany all future trial and guideline articles, as recommended by the ICMJE.[36] Guidance for mandatory COI disclosure items must be made clear for authors and enforced by the journal because authors may perceive differently what is reportable, allowing for bias.[47] Synchronizing standards among all publications would strengthen COI disclosure standards and the ability to review COIs between articles and databases. Public databases to capture COIs should be considered in countries beyond the United States and should include all scientists, rather than just physicians. Consideration should be made for the requirement of international authors to disclose COIs.

Using three publicly available databases ensured that conflicts in the articles were comprehensively analyzed. However, there are limitations to this study. First, there was an inability to identify COIs among international authors or nonphysician scientists, likely leading to underreporting of COIs. Second, the American Academy of Orthopaedic Surgeons Disclosure and ProPublica databases are self-reported by the authors, whereas the Open Payments database is updated by industry. There are generally more COI disclosures noted in the Open Payments database compared to the other two databases. This was primarily seen with COIs related to food and beverage, travel and lodging, education, and entertainment and would overestimate the rate of inaccuracy between COIs found in articles and those in databases. Third, it is possible that authors disclosed COIs, but the publishing entity did not include those disclosures in the publication. Lastly, the American Academy of Orthopaedic Surgeons Disclosure database primarily consists of orthopedic surgeons reducing surveillance of other medical specialties, many of which commonly use glucocorticoid injections.

Reporting of COIs is not fully adequate or transparent in current randomized controlled trials and clinical practice guidelines evaluating the effectiveness or recommendations for glucocorticoid injections in the management of knee osteoarthritis. This practice limits transparency and what can be known about the role of COIs within this body of research. COIs may impact trial results and clinical practice guideline recommendations, which is why the public should at least be aware of potential COIs when interpreting results and recommendations. Improvement with reporting practices and transparency with COI disclosures in future work are recommended.

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 Boyd 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. Brophy RH, Fillingham YA. AAOS Clinical practice guideline summary: management of osteoarthritis of the knee (non-arthroplasty), third edition. J Am Acad Orthop Surg 2022;30(9):e721–e729. doi:https://doi.org/10.5435/JAAOS-D-21-01233

2. Pavone V, Vescio A, Turchetta M, et al. Injection-based management of osteoarthritis of the knee: a systematic review of guidelines. Front Pharmacol 2021;12:661805. doi:https://doi.org/10.3389/fphar.2021.661805
3. 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. doi:https://doi.org/10.1002/art.41142
4. 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. doi:https://doi.org/10.1016/j.joca.2019.06.011
5. Royal Australian College of General Practitioners. Guideline for the management of knee and hip osteoarthritis: administrative report. 2nd ed. Royal Australian College of General Practitioners; 2018.
6. Rillo O, Riera H, Acosta C, et al. PANLAR consensus recommendations for the management in osteoarthritis of hand, hip, and knee. J Clin Rheumatol 2016;22(7):345–354. doi:https://doi.org/10.1097/RHU.0000000000000449
7. Berger A, Bozic K, Stacey B, et al. Patterns of pharmacotherapy and health care utilization and costs prior to total hip or total knee replacement in patients with osteoarthritis. Arthritis Rheum 2011;63(8):2268–2275. doi:https://doi.org/10.1002/art.30417
8. Dhawan A, Mather RC III, 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. doi:https://doi.org/10.1016/j.arthro.2013.09.002
9. Wijn SRW, Rovers MM, van Tienen TG, et al. Intra-articular corticosteroid injections increase the risk of requiring knee arthroplasty. Bone Joint J 2020;102-B(5):586–592. doi:https://doi.org/10.1302/0301-620X.102B5.BJJ-2019-1376.R1
10. Blankstein M, Lentine B, Nelms NJ. Common practices in intra-articular corticosteroid injection for the treatment of knee osteoarthritis: a survey of the American Association of Hip and Knee Surgeons membership. J Arthroplasty 2021;36(3):845–850. doi:https://doi.org/10.1016/j.arth.2020.09.022
11. Baker JF, Olave M, Leach W, et al. Corticosteroid injections for symptomatic treatment of osteoarthritis of the knee: a pilot blinded randomized trial. ACR Open Rheumatol 2023;5(10):529–535. doi:https://doi.org/10.1002/acr2.11596
12. Liu SH, Dubé CE, Eaton CB, et al. Longterm effectiveness of intraarticular injections on patient-reported symptoms in knee osteoarthritis. J Rheumatol 2018;45(9):1316–1324. doi:https://doi.org/10.3899/jrheum.171385
13. McAlindon TE, LaValley MP, Harvey WF, et al. Effect of intra-articular triamcinolone vs saline on knee cartilage volume and pain in patients with knee osteoarthritis: a randomized clinical trial. JAMA 2017;317(19):1967–1975. doi:https://doi.org/10.1001/jama.2017.5283
14. Kompel AJ, Roemer FW, Murakami AM, et al. Intra-articular corticosteroid injections in the hip and knee: perhaps not as safe as we thought? Radiology 2019;293(3):656–663. doi:https://doi.org/10.1148/radiol.2019190341
15. Deyle GD, Allen CS, Allison SC, et al. Physical therapy versus glucocorticoid injection for osteoarthritis of the knee. N Engl J Med 2020;382(15):1420–1429. doi:https://doi.org/10.1056/NEJMoa1905877
16. Soriano-Maldonado A, Klokker L, Bartholdy C, et al. Intra-articular corticosteroids in addition to exercise for reducing pain sensitivity in knee osteoarthritis: exploratory outcome from a randomized controlled trial. PLoS One 2016;11(2):e0149168. doi:https://doi.org/10.1371/journal.pone.0149168
17. Henriksen M, Christensen R, Klokker L, et al. Evaluation of the benefit of corticosteroid injection before exercise therapy in patients with osteoarthritis of the knee: a randomized clinical trial. JAMA Intern Med 2015;175(6):923–930. doi:https://doi.org/10.1001/jamainternmed.2015.0461
18. Jüni P, Hari R, Rutjes AWS, et al. Intra-articular corticosteroid for knee osteoarthritis. Cochrane Database Syst Rev 2015;2015(10):CD005328. doi:https://doi.org/10.1002/14651858.CD005328.pub3
19. 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. doi:https://doi.org/10.1001/jamanetworkopen.2021.42709
20. Gibbs AJ, Gray B, Wallis JA, et al. Recommendations for the management of hip and knee osteoarthritis: a systematic review of clinical practice guidelines. Osteoarthritis Cartilage 2023;31(10):1280–1292. doi:https://doi.org/10.1016/j.joca.2023.05.015
21. National Institute for Health and Care Excellence. Osteoarthritis in over 16s: Diagnosis and Management. National Institute for Health and Care Excellence; 2022. NICE Guideline, No. 226. Accessed November 4, 2025. https://pubmed.ncbi.nlm.nih.gov/36745715/
22. Romain PL. Conflicts of interest in research: looking out for number one means keeping the primary interest front and center. Curr Rev Musculoskelet Med 2015;8(2):122–127. doi:https://doi.org/10.1007/s12178-015-9270-2
23. Johnson J, Rogers W. Joint issues – conflicts of interest, the ASR hip and suggestions for managing surgical conflicts of interest. BMC Med Ethics 2014;15(1):63. doi:https://doi.org/10.1186/1472-6939-15-63
24. Weinfurt KP, Friedman JY, Dinan MA, et al. Disclosing conflicts of interest in clinical research: views of institutional review boards, conflict of interest committees, and investigators. J Law Med Ethics 2006;34(3):581–591. doi:https://doi.org/10.1111/j.1748-720X.2006.00072.x
25. Lin DH, Lucas E, Murimi IB, et al. Financial conflicts of interest and the centers for disease control and prevention's 2016 guideline for prescribing opioids for chronic pain. JAMA Intern Med 2017;177(3):427–428. doi:https://doi.org/10.1001/jamainternmed.2016.8471
26. Spithoff S, Leece P, Sullivan F, et al. Drivers of the opioid crisis: an appraisal of financial conflicts of interest in clinical practice guideline panels at the peak of opioid prescribing. PLoS One 2020;15(1):e0227045. doi:https://doi.org/10.1371/journal.pone.0227045
27. Institute of Medicine (US) Committee on Conflict of Interest. Conflicts of interest and development of clinical practice guidelines. In: Lo B, Field MJ, eds. Conflict of Interest in Medical Research, Education, and Practice. National Academies Press; 2009, 189–215.
28. DeFrance MJ, Yayac MF, Courtney PM, et al. The impact of author financial conflicts on robotic-assisted joint arthroplasty research. J Arthroplasty 2021;36(4):1462–1469. doi:https://doi.org/10.1016/j.arth.2020.10.033
29. Narain AS, Hijji FY, Yom KH, et al. Cervical disc arthroplasty: do conflicts of interest influence the outcome of clinical studies? Spine J 2017;17(7):1026–1032. doi:https://doi.org/10.1016/j.spinee.2017.03.018
30. Choo KJ, Yi PH, Burns R, et al. Variable reporting by authors presenting arthroplasty research at multiple annual conferences. J Arthroplasty 2017;32(1):315–319. doi:https://doi.org/10.1016/j.arth.2016.05.068
31. Ehrhardt S, Appel LJ, Meinert CL. Trends in National Institutes of Health funding for clinical trials registered in ClinicalTrials.gov. JAMA 2015;314(23):2566–2567. doi:https://doi.org/10.1001/jama.2015.12206
32. DeRosa GP, Angelos P, Barrack RL, et al. Report from the Task Force on Surgeon-Industry Relationships in the discipline of orthopaedic surgery. American Orthopaedic Association Orthopaedic Institute of Medicine; 2012. Accessed November 1, 2025. https://cdn-links.lww.com/permalink/jbjs/b/jbjs_2017_02_26_aoaoim_e89_sdc1.pdf

33. Lundh A, Lexchin J, Mintzes B, et al. Industry sponsorship and research outcome. Cochrane Database Syst Rev 2017;2017(2):MR000033. doi:https://doi.org/10.1002/14651858.MR000033.pub3
34. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467–473. doi:https://doi.org/10.7326/M18-0850
35. Covidence Systematic Review Software. Veritas Health Innovation; 2024. Accessed January 22, 2024. http://www.covidence.org
36. Drazen JM, Van Der Weyden MB, Sahni P, et al. Uniform format for disclosure of competing interests in ICMJE journals. N Engl J Med 2009; 361(19):1896–1897. doi:https://doi.org/10.1056/NEJMe0909052
37. Rosenberg J, Bauchner H, Backus J, et al. The new ICMJE recommendations. Dan Med J 2013;60(10):A4718.
38. Tisherman RT, Wawrose RA, Mittal AM, et al. Conflict of interest disclosure in American arthroplasty surgical literature. Arthroplast Today 2024;30:101493. doi:https://doi.org/10.1016/j.artd.2024.101493
39. Okike K, Kocher MS, Wei EX, et al. Accuracy of conflict-of-interest disclosures reported by physicians. N Engl J Med 2009;361(15): 1466–1474. doi:https://doi.org/10.1056/NEJMsa0807160
40. Buerba RA, Fu MC, Grauer JN. Discrepancies in spine surgeon conflict of interest disclosures between a national meeting and physician payment listings on device manufacturer web sites. Spine J 2013;13(12): 1780–1788. doi:https://doi.org/10.1016/j.spinee.2013.05.032
41. Hughes JD, Shin JJ, Albers M, et al. A closer look at the relationship between industry and orthopaedic sports medicine surgeons. Orthop J Sports Med 2019;7(1):2325967118823175. doi:https://doi.org/10.1177/2325967118823175
42. Ju B, Miller C, Whang P, et al. Comparing financial disclosure reporting at annual spine conferences [abstract]. Spine J 2009;9(10):49S. doi:https://doi.org/10.1016/j.spinee.2009.08.117
43. Mayo BC, Ravella KC, Onsen L, et al. Is there an association between authors' conflicts of interest and outcomes in clinical studies involving autologous chondrocyte implantation? Orthop J Sports Med 2021;9(2):2325967120979988. doi:https://doi.org/10.1177/2325967120979988
44. Chimonas S, Frosch Z, Rothman DJ. From disclosure to transparency: the use of company payment data. Arch Intern Med 2011; 171(1):81–86. doi:https://doi.org/10.1001/archinternmed.2010.341
45. Flanagin A, Fontanarosa PB, DeAngelis CD. Update on JAMA's conflict of interest policy. JAMA 2006;296(2):220–221. doi:https://doi.org/10.1001/jama.296.2.220
46. Dickerson DM, Naidu RK. Preparing for the physician payment sunshine act. Reg Anesth Pain Med 2014;39(3):185–188. doi:https://doi.org/10.1097/AAP.0000000000000074
47. McCoy MS, Emanuel EJ. Why there are no "potential" conflicts of interest. JAMA 2017;317(17):1721–1722. doi:https://doi.org/10.1001/jama.2017.2308

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1209–1219
DOI 10.1002/acr.70040
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

American College of Rheumatology
Empowering Rheumatology Professionals

Associations of Sleep and Shift Work With Osteoarthritis Risk

Elizabeth L. Yanik, PhD, ScM,[1] Abigail Bridgeman,[1] Erik D. Herzog, PhD,[1] Vy Pham, MPH,[1] Bradley A. Evanoff, MD, MPH,[1] and Farshid Guilak, PhD[1,2]

Objective. Daily rhythms may be critical for maintaining homeostasis of joint tissues. We aimed to investigate the relationships among circadian clock disruption, sleep, and osteoarthritis (OA) risk in humans.

Methods. In the UK Biobank, a prospective 500,000–person cohort, we evaluated associations among sleep duration, sleeplessness/insomnia, and shift work type with four endpoints: knee OA, hip OA, total knee arthroplasty (TKA), and total hip arthroplasty. Cox regression was used to estimate associations with OA endpoints adjusting for age, sex, education, race, Townsend Deprivation Index, manual work frequency, and frequency of occupational walking/standing. Associations with and without adjustment for body mass index were estimated, as circadian clock disruption may influence OA through effects on obesity.

Results. For all OA endpoints, risk was highest among those getting <6 hours of nightly sleep (eg, hazard ratios [HRs] for <6 vs 7 hours: 1.21–1.41), and "usually" experiencing sleeplessness/insomnia compared with "never/rarely" was associated with higher risk (HRs: 1.24–1.40). Night shift workers had 24% higher knee OA risk (HR = 1.24; 95% confidence interval [CI] = 1.12–1.38) and 28% higher TKA risk (HR = 1.28; 95% CI = 1.19–1.37) compared with nonshift workers. After controlling for body mass index, associations were attenuated, but short sleep and sleeplessness/insomnia remained associated with all endpoints, and night shift work remained associated with knee OA and TKA. Sleep associations were similar after excluding participants reporting chronic knee/hip pain at sleep assessment.

Conclusion. Disruption of sleep or circadian rhythms may be modifiable risk factors for OA underlying cartilage degeneration through obesity and obesity-independent pathways. These findings point to potential ways to prevent OA.

INTRODUCTION

Osteoarthritis (OA) is a multifactorial degenerative joint disease, with primary risk factors that include joint injury, metabolic dysfunction, aging, and genetics.[1] But the mechanisms through which these factors influence OA have not been fully elucidated. Recent evidence suggests that circadian rhythm disruptions may be a key mechanism in OA.[2] Daily rhythms in physiology and behavior depend, partly, on circadian neurons and glia in the suprachiasmatic nucleus of the hypothalamus that coordinate circadian cells in other tissues.[3] These circadian rhythms can also be influenced by external cues, such as light exposure, diet, and physical activity.[4] Circadian rhythms are crucial for maintaining homeostasis of the articular cartilage and other joint tissues.[5,6] Cartilage undergoes daily cycles of high load and metabolic demand, followed by periods of low load that allow recovery of tissue thickness and water content. Chondrocytes, the sole cell population in cartilage, exhibit cell-autonomous circadian rhythms in gene expression that can synchronize to local time through daily cycles of mechanical stimulation.[5,7,8] Furthermore, in rodent models, disruption of the circadian clock can lead to OA-like cartilage degeneration,[9–11] and changes in clock gene expression have been observed in OA cartilage.[12,13] Circadian clock disruptions also increase obesity,[3] a key OA risk factor,[1] and poor sleep associated with circadian clock disruption may influence pain perception.[14,15] Importantly, circadian clock disruptions are

Supported in part by the NIH (grants R03-AG-088746, R21-AR-083662, and R01-AR-083937) and by Shriners Hospitals for Children.

This research has been conducted using the UK Biobank Resource (https://www.ukbiobank.ac.uk/) under Application 27034. This work uses data provided by patients and collected by the NHS as part of their care and support and data assets made available by National Safe Haven as part of the Data and Connectivity National Core Study Programme.

[1]Washington University, St. Louis, Missouri; [2]Shriners Hospitals for Children, St. Louis, Missouri.

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.70040).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.70040.

Address correspondence via email to Elizabeth L. Yanik, PhD, ScM, at yanike@wustl.edu.

Submitted for publication July 14, 2025; accepted in revised form December 8, 2025.

SIGNIFICANCE & INNOVATIONS

- Our study leveraged the large, population-based UK Biobank cohort to demonstrate that short sleep duration, poor sleep quality, and night shift work were associated with higher risk of an incident osteoarthritis (OA) diagnosis and total joint arthroplasty, after controlling for physical work demands, socioeconomic status, and demographics.
- Poor sleep and shift work may partly influence OA through effects on weight gain, but significant associations with OA risk were still observed after controlling for body mass index (BMI), indicating BMI-independent mechanisms that affect OA.
- Short sleep and poor sleep had similar effects on knee and hip OA endpoints, whereas night shift work was only associated with knee OA, potentially indicating that circadian clock disruptions are particularly impactful for the knee.
- As circadian rhythms and sleep can be optimized through lifestyle changes and sleep hygiene interventions, these findings point to new potential ways to prevent OA development and progression.

potentially modifiable with interventions, such as light exposure therapy and sleep hygiene counseling.[16–19]

However, little is known about the relationship between circadian clock disruption and OA in human populations. Sleep patterns, including duration and quality, may indicate circadian clock disruption or other sleep disorders. Although sleep disruptions are known to occur in patients with prevalent OA,[20,21] few studies have investigated the effect of sleep disturbances occurring before symptomatic OA. Additionally, a better understanding of the role of environmental factors known to cause circadian clock disruption can point to potential prevention opportunities. Shift work, particularly night shift work, is a common environmental cause of circadian clock disruption among working adults. One study of Chinese automotive workers found that shift work was associated with 19% higher knee OA risk.[22] The role of chronotype, that is, the preference for morning or evening behavior, also has not been investigated as an OA risk factor. Those with an "eveningness" chronotype (preference for evening activity) may be unaffected or even benefit from night shift work, whereas night shift work could be most detrimental to those with a "morningness" chronotype (preference for morning activity). Studies have shown more sleep loss and sleep disturbance during night shifts in those with "morningness" chronotypes compared with "eveningness" chronotypes, whereas an inverse pattern is observed during early morning shifts.[23,24]

To provide insights into the relationships among sleep, circadian disruption, and OA, we evaluated associations of shift work, chronotype, sleep duration, and sleep quality with OA risk in the UK Biobank (UKB). A cross-sectional UKB analysis of occupational factors associated with OA found associations between shift work and OA but did not exclude prevalent OA cases in which symptoms and disability may have influenced work.[25] We hypothesized that short sleep duration, more frequent sleeplessness/insomnia, and shift work would be associated with higher rates of incident knee and hip OA diagnoses as well as total knee arthroplasty (TKA) and total hip arthroplasty (THA). This investigation serves as an important next step for understanding the role of circadian clock disruption in the pathobiology of OA in human populations, which could lead to new methods of preventing OA development and progression.

MATERIALS AND METHODS

We conducted a cohort study using the UKB, a population-based cohort of approximately 500,000 people in the United Kingdom. The UKB recruited people aged 40 to 69 years through 22 assessment centers between 2006 and 2010. After informed consent, enrolled participants completed detailed baseline questionnaires (covering demographics, occupational information, and health behaviors) and had physical measurements taken, including weight and height to calculate body mass index (BMI). A baseline interview was conducted with a trained nurse, collecting information on medical history (including self-reported OA, TKA, and THA). Participants can withdraw at any time (to date <0.01% have withdrawn).

In the present study, people were excluded if they had withdrawn from the UKB or had a record of the OA endpoint before enrollment (TKA, THA, knee OA, or hip OA). Individuals with inflammatory arthritis (eg, rheumatoid arthritis or psoriatic arthritis) or autoimmune disorders affecting the connective tissue (eg, lupus) were excluded from the study population for all analyses. Exclusions were made at baseline for those self-reporting inflammatory arthritis/autoimmune disorders at baseline or at time of diagnosis for those with a diagnosis in linked hospital records.

The UK NHS Research Ethics Committee approved the UKB. The Washington University Institutional Review Board determined that this study did not require ethical approval.

Sleep and occupational measures. The UKB baseline questionnaire asked about sleep, including sleep duration ("About how many hours sleep do you get in every 24 hours?"), sleeplessness/insomnia ("Usually/sometimes/rarely/never have trouble falling asleep at night or wake up in the middle of the night"), and chronotype ("Do you consider yourself 'definitely a morning person,' 'more a morning than evening person,' 'more an evening than a morning person,' or 'definitely an evening person'?"). The chronotype question is the question that explains the most variance in sleep-wake timing preferences in the validated Morningness-Eveningness Questionnaire.[26,27] We considered all sleep measures as categorical variables, including sleep duration (<6, 6, 7, 8, or >8 hours), as worse health

is often observed in both short and long sleepers. The questionnaire also included questions about work. Participants who reported paid employment or self-employment were asked about shift work frequency ("Work always/usually/sometimes/rarely/never involves 'shift work,' defined as work outside of 9 AM–5 PM work hours, or 'night shift work,' defined as work through hours from 12 AM–6 AM") and physical job demands ("Job always/usually/sometimes/rarely involves 'heavy manual or physical work' or 'mainly walking or standing'"). For the current study, people reporting "never/rarely" doing shift work were categorized as no shift work; people who reported "sometimes," "usually," or "always" doing night shift work were categorized as night shift work; and people who reported "sometimes," "usually," or "always" doing shift work but "never/rarely" doing night shift work were categorized as day shifts only.

OA and joint replacement. UK NHS hospital records are linked to the UKB cohort providing longitudinal data on hospital diagnoses and procedures from the baseline assessment visit until October 31, 2022, at the time of this study. Hospital diagnoses are coded with the *International Classification of Diseases, Tenth Revision* (*ICD-10*) and procedures are coded with the Office of Population Censuses and Surveys Classification, Fourth Revision (OPCS-4). Primary care records were linked to 163,000 participants, allowing OA diagnosis identification from both hospitals and primary care up until March 31, 2017, at the time of this study. Primary care diagnoses are coded with Read codes that map to *ICD-9/10* codes. Our study focused on four outcomes: incident hip OA, knee OA, THA, and TKA.

For incident hip/knee OA analyses, we included people with ≥36 months of active primary care registration and excluded people with prevalent OA determined by (1) OA diagnosis before 36 months of active registration or their baseline UKB visit or (2) self-reported prevalent OA at baseline. Incident hip or knee OA was defined as the presence of at least two primary care hip/knee OA Read codes within 6 months or a hospital *ICD-10* code for hip/knee OA.[28,29]

Primary TKA and THA procedures were identified with OPCS-4 codes used by the UK National Joint Registry. As not all TKAs/THAs are attributable to OA, we only included cases when the participant had self-reported OA or had an OA diagnosis by the time of the procedure. We excluded TKA/THAs that occurred concurrently with a fracture diagnosis, an indication that the procedure may not be attributable to OA.

Follow-up time started at the baseline UKB visit. For incident hip/knee OA endpoints, follow-up ended at the first of incident hip/knee OA, death, or the last date of available primary care records. For THA/TKA endpoints, follow-up ended at the first of THA/TKA date, death, or the last date of available hospital records.

Statistical analyses. Analyses of sleep duration, sleeplessness/insomnia, and chronotype were limited to participants who responded to the relevant questions at their baseline visit (>99% of population for sleep duration and sleeplessness/insomnia and 89% of population for chronotype). Shift work analyses included the UKB working population reporting shift work frequency.

Multivariable Cox regression models were used to estimate associations with each OA endpoint using SAS (version 9.4). The UKB collected information on numerous factors included in models as potential confounders: age, sex, ethnicity, heavy physical work frequency, occupational walking/standing frequency, Townsend Deprivation Index (TDI, a measure of neighborhood socioeconomic status), and education. All associations were considered with and without adjustment for BMI, as circadian clock disruption may influence OA through effects on obesity. BMI may also serve as a confounder of sleep associations, as obesity can lead to worse sleep, particularly through obesity-associated sleep apnea.[30]

Sleep associations with OA endpoints could be influenced by symptomatic, undiagnosed OA. To avoid this potential bias, we conducted sensitivity analyses that excluded participants reporting ≥3 months of knee pain (for knee OA and TKA associations), or ≥3 months of hip pain (for hip OA and THA associations) at their baseline visit.

We evaluated the modification of shift work effects by chronotype and physical work frequency and modification of sleep duration effects by shift work type. Effect modification was evaluated quantitatively through interaction terms in Cox regression models adjusting for the same covariates included in the main analysis ($P < 0.10$ used to identify significant interactions) and qualitatively through graphing results of stratified analyses. If the largest relative associations were observed in the strata with lowest OA risk, we further calculated a relative excess risk for interaction (RERI) statistic to evaluate whether effect modification was present on the absolute risk scale.[31]

Data availability statement. Data sets supporting the conclusions of this article are available in the UKB repository, (https://www.ukbiobank.ac.uk/). All bona fide researchers can apply to access the UKB repository to conduct health-related research in the public interest. Completion of a Material Transfer Agreement is required for data access.

RESULTS

Population characteristics. Among the 502,363 UKB participants, 54% were female, 67% had an overweight or obese BMI, and the median age was 58 years at the baseline visit when shift work and sleep measures were collected (Supplementary Table S1). Median sleep duration was seven hours; 48% reported "sometimes" experiencing sleeplessness/insomnia, and 28%

reported "usually" experiencing sleeplessness/insomnia (<1% participants did not report their sleeplessness/insomnia frequency). For chronotype, 62% of participants reported being "definitely a morning person" or "more a morning than evening person" (<12% of participants did not report chronotype; Supplementary Table S1).

At the baseline UKB visit, 13,720 participants reported prevalent OA, 6,038 had a history of TKA, and 7,650 had a THA history. These participants were excluded from analyses for the corresponding endpoints. TKA and THA endpoints were assessed over a mean 13 follow-up years in hospital records. Knee and hip OA endpoints were assessed over a mean 8 follow-up years in primary care records.

Sleep duration, sleeplessness, and chronotype associations. People reporting more frequent sleeplessness/insomnia were more likely to be older, female, White, and obese and have less than a college education (Supplementary Table S2). Shorter sleep duration and more frequent sleeplessness/insomnia were associated with higher risk of all OA endpoints even after adjusting for age, sex, race, education, TDI, manual/physical work frequency, and occupational walking/standing frequency (Table 1). People getting <6 hours of sleep daily had 41% higher knee OA risk (hazard ratio [HR] = 1.41, 95% confidence interval [CI] = 1.29–1.53) and 31% higher TKA risk (HR = 1.31, 95% CI = 1.24–1.38) compared with people who slept 7 hours daily. People that reported usually having trouble falling asleep or waking up in the night had 34% higher knee OA risk (HR = 1.34, 95% CI = 1.26–1.43) and 40% higher TKA risk (HR = 1.40, 95% CI = 1.34–1.46) compared with people who never/rarely had these problems. Similar elevations in risk associated with short sleep and sleeplessness were observed for hip OA and THA (Table 1).

People who considered themselves to be "definitely a morning person" had higher knee OA and TKA risk than any other chronotype (Table 1). By contrast, hip OA and THA risk did not differ significantly across the chronotype groups.

Table 1. Associations of sleep and shift work with OA endpoints among UKB participants*

	Knee OA		TKA		Hip OA		THA	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
Sleep duration								
<6 h	1.41 (1.29–1.53)	<0.01	1.31 (1.24–1.38)	<0.01	1.39 (1.24–1.56)	<0.01	1.21 (1.13–1.28)	<0.01
6 h	1.19 (1.12–1.27)	<0.01	1.17 (1.13–1.22)	<0.01	1.17 (1.08–1.27)	<0.01	1.11 (1.06–1.15)	<0.01
7 h	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
8 h	0.95 (0.89–1.00)	0.05	0.96 (0.92–0.99)	0.01	1.02 (0.95–1.09)	0.68	0.99 (0.95–1.03)	0.62
>8 h	0.98 (0.90–1.07)	0.72	1.04 (0.98–1.10)	0.17	0.99 (0.88–1.10)	0.78	1.05 (1.00–1.12)	0.07
Sleeplessness/insomnia								
Never/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.09 (1.03–1.16)	<0.01	1.13 (1.08–1.17)	<0.01	1.09 (1.01–1.17)	0.03	1.06 (1.02–1.10)	<0.01
Usually	1.34 (1.26–1.43)	<0.01	1.40 (1.34–1.46)	<0.01	1.30 (1.20–1.41)	<0.01	1.24 (1.19–1.30)	<0.01
Chronotype								
Definitely a morning person	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
More a morning than evening person	0.89 (0.84–0.94)	<0.01	0.90 (0.87–0.93)	<0.01	0.96 (0.89–1.03)	0.27	1.00 (0.96–1.04)	0.93
More an evening than morning person	0.90 (0.84–0.96)	<0.01	0.90 (0.86–0.93)	<0.01	0.98 (0.91–1.07)	0.69	0.98 (0.94–1.03)	0.45
Definitely an evening person	0.95 (0.87–1.05)	0.32	0.92 (0.87–0.98)	0.01	1.01 (0.90–1.14)	0.84	1.02 (0.96–1.09)	0.53
Shift work frequency								
None/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.17 (1.05–1.31)	0.01	1.24 (1.15–1.34)	<0.01	0.96 (0.81–1.14)	0.63	0.96 (0.88–1.05)	0.41
Usually	0.98 (0.80–1.21)	0.88	1.11 (0.97–1.27)	0.14	0.98 (0.73–1.31)	0.86	1.01 (0.86–1.18)	0.94
Always	1.20 (1.08–1.34)	<0.01	1.21 (1.12–1.30)	<0.01	1.13 (0.96–1.33)	0.15	1.01 (0.92–1.10)	0.90
Night shift frequency								
None/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.32 (1.16–1.50)	<0.01	1.30 (1.19–1.42)	<0.01	0.98 (0.79–1.21)	0.85	0.97 (0.87–1.09)	0.62
Usually	1.02 (0.79–1.32)	0.90	1.13 (0.94–1.35)	0.18	0.78 (0.50–1.20)	0.25	0.98 (0.80–1.21)	0.87
Always	1.17 (0.98–1.39)	0.09	1.23 (1.09–1.39)	<0.01	1.17 (0.90–1.52)	0.24	0.97 (0.83–1.13)	0.66
Type of shift work								
None	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Only day shifts	1.07 (0.96–1.20)	0.20	1.14 (1.06–1.22)	<0.01	1.06 (0.91–1.23)	0.46	1.01 (0.93–1.09)	0.91
Night shifts	1.24 (1.12–1.38)	<0.01	1.28 (1.19–1.37)	<0.01	1.01 (0.86–1.19)	0.89	0.97 (0.89–1.06)	0.53

* All models adjusted for age, sex, education, race, Townsend Deprivation Index, frequency of manual/physical labor, and frequency of occupational walking/standing. Missing indicator variables were included when data were missing for race, education, frequency of manual/physical labor, and frequency of occupational walking/standing. For sleep duration, sleeplessness/insomnia, and chronotype, the analysis included 144,600 UKB participants for knee OA (with 7,996 cases), 145,658 for hip OA (with 4,969 cases), 489,340 for TKA (with 19,888 cases) and 486,242 for THA (with 17,414 cases). For shift work, the analysis included 84,607 UKB workers for knee OA (with 3,877 cases), 282,630 UKB workers for TKA (with 8,336 cases), 85,147 UKB workers for hip OA (with 2,202 cases), and 281,412 UKB workers for THA (with 7,718 cases). CI, confidence interval; HR, hazard ratio; OA, osteoarthritis; Ref., reference; THA, total hip arthroplasty; TKA, total knee arthroplasty; UKB, UK Biobank.

Table 2. Body mass index–independent associations of sleep and shift work with OA endpoints among UKB participants*

	Knee OA		TKA		Hip OA		THA	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
Sleep duration								
<6 h	1.31 (1.20–1.43)	<0.01	1.20 (1.14–1.27)	<0.01	1.35 (1.20–1.51)	<0.01	1.17 (1.10–1.25)	<0.01
6 h	1.15 (1.08–1.22)	<0.01	1.12 (1.08–1.17)	<0.01	1.16 (1.07–1.25)	<0.01	1.09 (1.04–1.13)	<0.01
7 h	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
8 h	0.94 (0.89–1.00)	0.04	0.95 (0.92–0.98)	<0.01	1.01 (0.94–1.09)	0.75	0.99 (0.95–1.03)	0.51
>8 h	0.94 (0.86–1.02)	0.13	0.95 (0.90–1.01)	0.08	0.96 (0.86–1.07)	0.48	1.02 (0.96–1.08)	0.47
Sleeplessness/insomnia								
Never/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.09 (1.02–1.15)	0.01	1.12 (1.08–1.17)	<0.01	1.09 (1.01–1.17)	0.03	1.05 (1.01–1.10)	0.01
Usually	1.29 (1.21–1.38)	<0.01	1.33 (1.28–1.39)	<0.01	1.28 (1.18–1.39)	<0.01	1.22 (1.17–1.27)	<0.01
Chronotype								
Definitely a morning person	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
More a morning than evening person	0.91 (0.86–0.96)	<0.01	0.92 (0.89–0.96)	<0.01	0.97 (0.90–1.04)	0.41	1.01 (0.97–1.05)	0.67
More an evening than morning person	0.89 (0.84–0.95)	<0.01	0.89 (0.86–0.93)	<0.01	0.98 (0.90–1.06)	0.59	0.98 (0.94–1.02)	0.38
Definitely an evening person	0.91 (0.83–1.00)	0.05	0.86 (0.81–0.92)	<0.01	0.99 (0.88–1.11)	0.83	1.00 (0.93–1.06)	0.86
Shift work frequency								
None/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.10 (0.98–1.23)	0.10	1.15 (1.06–1.24)	<0.01	0.93 (0.79–1.10)	0.39	0.94 (0.86–1.03)	0.16
Usually	0.96 (0.78–1.18)	0.69	1.04 (0.91–1.20)	0.54	0.96 (0.71–1.29)	0.78	0.99 (0.84–1.15)	0.85
Always	1.12 (1.01–1.25)	0.04	1.10 (1.02–1.19)	0.02	1.09 (0.93–1.28)	0.30	0.97 (0.89–1.07)	0.54
Night shift frequency								
None/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.24 (1.09–1.40)	<0.01	1.19 (1.09–1.31)	<0.01	0.95 (0.77–1.17)	0.61	0.94 (0.84–1.06)	0.30
Usually	0.96 (0.74–1.24)	0.76	1.03 (0.87–1.23)	0.72	0.75 (0.49–1.16)	0.20	0.95 (0.77–1.18)	0.64
Always	1.09 (0.91–1.31)	0.33	1.12 (0.99–1.27)	0.07	1.13 (0.87–1.47)	0.37	0.93 (0.80–1.09)	0.35
Type of shift work								
None	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Only day shifts	1.03 (0.92–1.14)	0.66	1.07 (0.99–1.15)	0.08	1.03 (0.89–1.20)	0.69	0.98 (0.91–1.07)	0.67
Night shifts	1.15 (1.04–1.28)	0.01	1.16 (1.08–1.25)	<0.01	0.97 (0.83–1.15)	0.74	0.94 (0.86–1.02)	0.15

* All models adjusted for age, sex, education, race, Townsend Deprivation Index, frequency of manual/physical labor, frequency of occupational walking/standing, and body mass index. Missing indicator variables were included when data were missing for race, education, frequency of manual/physical labor, frequency of occupational walking/standing, and body mass index. For sleep duration, sleeplessness/insomnia, and chronotype, the analysis included 144,600 UKB participants for knee OA (with 7,996 cases), 145,658 for hip OA (with 4,969 cases), 489,340 for TKA (with 19,888 cases) and 486,242 for THA (with 17,414 cases). For shift work, the analysis included 84,607 UKB workers for knee OA (with 3,877 cases), 282,630 UKB workers for TKA (with 8,336 cases), 85,147 UKB workers for hip OA (with 2,202 cases), and 281,412 UKB workers for THA (with 7,718 cases). CI, confidence interval; HR, hazard ratio; OA, osteoarthritis; Ref., reference; THA, total hip arthroplasty; TKA, total knee arthroplasty; UKB, UK Biobank.

After additional BMI adjustment, similar results were observed for sleep duration, sleeplessness, and chronotype, although associations were attenuated (Table 2). For instance, people getting <6 hours of sleep daily had 31% higher knee OA risk (HR = 1.31) and 20% higher TKA risk (HR = 1.20) compared with people who slept 7 hours daily after adjusting for BMI.

Prevalent knee or hip pain in participants with undiagnosed OA could have influenced sleep associations. Thus, we conducted sensitivity analyses in which participants who reported chronic (≥3 months) knee pain were excluded from knee OA/TKA analyses, and participants who reported chronic hip pain were excluded from hip OA/THA analyses. Results were similar after these exclusions (Table 3). For example, in people without chronic knee/hip pain, getting <6 hours of sleep daily was associated with 37% higher knee OA risk (HR = 1.37), 25% higher TKA risk (HR = 1.25), 37% higher hip OA risk (HR = 1.37), and 15% higher THA risk (HR = 1.15; Table 3).

Shift work associations. There were 215,316 UKB participants unemployed or retired at the time of their baseline assessment. Among the 287,047 working participants included in shift work analyses, 9% reported night shift work, 8% reported shift work only during the day, 7% reported jobs that always required heavy manual/physical work, and 19% reported jobs that always involved mainly walking or standing (Supplementary Table S3). Shift workers were more likely to be male, non-White, and have less than a college education (Supplementary Table S4). Shift workers also reported more frequent heavy physical work and were more likely to have an obese BMI.

Compared with workers who did not work shifts, day shift workers had 7% higher knee OA risk (HR = 1.07, 95%

Table 3. Associations of sleep with OA endpoints among UK Biobank participants without chronic hip/knee pain at baseline*

	Knee OA		TKA		Hip OA		THA	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
BMI unadjusted								
Sleep duration								
<6 h	1.37 (1.22–1.54)	<0.01	1.25 (1.14–1.36)	<0.01	1.37 (1.20–1.56)	<0.01	1.15 (1.07–1.24)	<0.01
6 h	1.18 (1.09–1.27)	<0.01	1.13 (1.06–1.19)	<0.01	1.16 (1.06–1.26)	<0.01	1.08 (1.02–1.13)	<0.01
7 h	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
8 h	0.95 (0.89–1.03)	0.20	0.94 (0.90–1.00)	0.03	0.99 (0.91–1.07)	0.77	0.98 (0.94–1.02)	0.31
>8 h	0.98 (0.88–1.09)	0.70	1.02 (0.94–1.11)	0.58	0.95 (0.84–1.08)	0.43	1.04 (0.97–1.12)	0.23
Sleeplessness/insomnia								
Never/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.06 (0.99–1.14)	0.10	1.09 (1.03–1.15)	<0.01	1.07 (0.98–1.17)	0.12	1.03 (0.98–1.08)	0.22
Usually	1.21 (1.11–1.31)	<0.01	1.26 (1.18–1.34)	<0.01	1.26 (1.15–1.38)	<0.01	1.13 (1.08–1.19)	<0.01
Chronotype								
Definitely a morning person	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
More a morning than evening person	0.88 (0.82–0.95)	<0.01	0.87 (0.83–0.92)	<0.01	0.94 (0.87–1.02)	0.16	1.00 (0.95–1.04)	0.85
More an evening than morning person	0.91 (0.84–0.98)	0.02	0.87 (0.83–0.93)	<0.01	0.95 (0.87–1.04)	0.25	0.97 (0.93–1.02)	0.29
Definitely an evening person	0.90 (0.80–1.02)	0.10	0.82 (0.75–0.90)	<0.01	0.96 (0.83–1.10)	0.55	1.03 (0.95–1.11)	0.52
BMI adjusted								
Sleep duration								
<6 h	1.29 (1.15–1.45)	<0.01	1.16 (1.06–1.26)	<0.01	1.33 (1.17–1.52)	<0.01	1.12 (1.04–1.21)	<0.01
6 h	1.14 (1.06–1.24)	<0.01	1.09 (1.02–1.15)	0.01	1.14 (1.04–1.25)	<0.01	1.06 (1.01–1.11)	0.03
7 h	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
8 h	0.95 (0.89–1.02)	0.17	0.94 (0.89–0.99)	0.02	0.98 (0.91–1.07)	0.69	0.97 (0.93–1.02)	0.24
>8 h	0.94 (0.84–1.05)	0.26	0.95 (0.87–1.02)	0.17	0.93 (0.82–1.05)	0.25	1.01 (0.95–1.09)	0.72
Sleeplessness/insomnia								
Never/rarely	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sometimes	1.06 (0.99–1.14)	0.10	1.09 (1.03–1.15)	<0.01	1.07 (0.98–1.16)	0.12	1.03 (0.98–1.08)	0.25
Usually	1.18 (1.09–1.28)	<0.01	1.22 (1.15–1.29)	<0.01	1.24 (1.13–1.36)	<0.01	1.11 (1.06–1.17)	<0.01
Chronotype								
Definitely a morning person	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
More a morning than evening person	0.90 (0.84–0.97)	0.01	0.70 (0.85–0.95)	<0.01	0.95 (0.88–1.04)	0.26	1.01 (0.96–1.06)	0.78
More an evening than morning person	0.90 (0.83–0.98)	0.01	0.87 (0.82–0.93)	<0.01	0.94 (0.86–1.03)	0.20	0.97 (0.92–1.02)	0.26
Definitely an evening person	0.87 (0.77–0.98)	0.02	0.77 (0.70–0.85)	<0.01	0.94 (0.82–1.08)	0.36	1.00 (0.93–1.08)	0.99

* All models adjusted for age, sex, education, race, Townsend Deprivation Index, frequency of manual/physical labor, and frequency of occupational walking/standing. Missing indicator variables were included when data was missing for race, education, frequency of manual/physical labor, and frequency of occupational walking/standing. After excluding people with chronic knee pain, there were 124,769 people and 4,825 knee OA diagnoses in the analytic population for knee OA and 411,165 people and 8,915 TKA cases in the analytic population for TKA. After excluding people with chronic hip pain there were 135,659 people and 3,797 hip OA diagnoses in the analytic population for hip OA and 447,405 people and 12,298 THA cases in the analytic population for THA. BMI, body mass index; CI, confidence interval; HR, hazard ratio; OA, osteoarthritis; Ref., reference; THA, total hip arthroplasty; TKA, total knee arthroplasty.

CI = 0.96–1.20; $P = 0.20$) and 14% higher TKA risk (HR = 1.14, 95% CI = 1.06–1.22; $P < 0.01$) after adjusting for age, sex, race, education, TDI, manual/physical work frequency, and occupational walking/standing frequency (Table 1). In the same analyses, night shift workers had 24% higher knee OA risk (HR = 1.24, 95% CI = 1.12–1.38; $P < 0.01$) and 28% higher TKA risk (HR = 1.28 95% CI = 1.19–1.37; $P < 0.01$) compared with nonshift workers. After additional BMI adjustment, all associations were attenuated, but night shift work remained significantly associated with both knee OA and TKA (Table 2). Evaluation of day shift frequency and night shift frequency (reported as work that "sometimes," "usually," or "always" involves shift work) did not indicate a consistent trend between increasing frequency of shift work and higher risk of knee OA or TKA (Table 1). Day shift and night shift work were not significantly associated with hip OA or THA, regardless of adjustment for BMI (Tables 1 and 2).

Interactions of shift work with chronotype, physical work, and sleep duration. Day shift work effects did not differ by chronotype (Figure 1). For night shift effects on TKA risk, there was an interaction with chronotype ($P < 0.10$) in which night shift effects were larger among people with a morning chronotype (Figure 1B and D). However, for knee OA risk there was no evidence of a larger effect of night shift work among people with a morning chronotype in terms of either a relative risk increase (Figure 1A and C) or absolute risk increase (RERI = –0.04 indicating that the combined effect of morning chronotype and night shift work on excess risk was 0.04 less than the sum of the individual

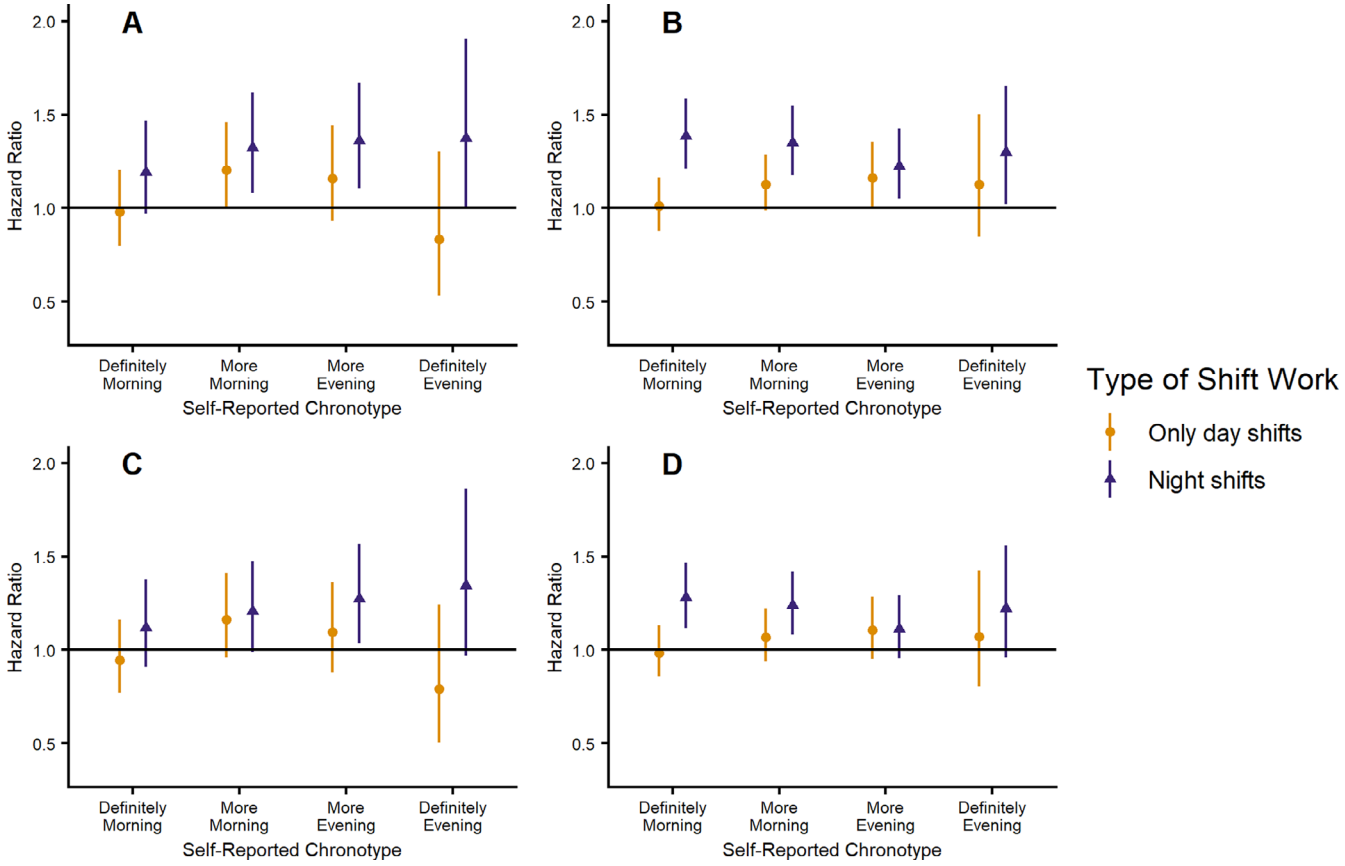

Figure 1. Shift work associations with knee osteoarthritis outcomes stratified by chronotype. (A) Associations with knee osteoarthritis, not adjusted for body mass index. (B) Associations with total knee arthroplasty, not adjusted for body mass index. (C) Associations with knee osteoarthritis, adjusted for body mass index. (D) Associations with total knee arthroplasty, adjusted for body mass index. All associations adjusted for age, sex, education, race, Townsend Deprivation Index, frequency of manual/physical labor, and frequency of occupational walking/standing. Nonshift workers are the reference group for each hazard ratio. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.70040/abstract.

effects). Chronotype interactions were not observed for hip OA or THA.

After stratifying by frequency of physical work, shift work effects on both knee OA and TKA were strongest among workers with jobs that never/rarely require heavy physical work (Figure 2). In these workers, day shift work was associated with 18% higher risk of knee OA and 26% higher risk of TKA, whereas night shift work was associated with 43% higher risk of knee OA and 40% higher risk of TKA (Figure 2A and B). As in the main analysis, adjustment for BMI attenuated associations, but shift work remained associated with elevated knee OA and TKA risk among workers with jobs that were not physically demanding (Figure 2C and D). By contrast, among workers with jobs that usually or always require heavy physical work, shift work was associated with smaller relative increases in TKA risk and smaller absolute increases in TKA risk (RERI = −1.31 and −1.62 for day shift work and night shift work, respectively, indicating that the combined effects of heavy physical work and shift work on excess risk were 1.31 and 1.62 less than the sum of the individual effects). Night shift work was associated with no increase in knee OA risk among workers with jobs that always require heavy physical work.

We also evaluated associations of sleep duration with OA endpoints in workers stratified by shift work type. Statistically significant associations between short sleep and OA endpoints were primarily seen among workers who did not do shift work (Figure 3). Association estimates between short sleep and OA endpoints in day and night shift workers were imprecise as a minority of workers do shift work, but HRs still indicated elevations in risk.

DISCUSSION

Our study leveraged the large, population-based UKB cohort to demonstrate that short sleep duration, more frequent sleeplessness/insomnia, and night shift work were associated with higher OA risk, after controlling for physical work demands, socioeconomic status, and demographics. Poor sleep and shift work may partly influence OA risk through effects on weight gain.[3,32] However, significant associations with OA risk were still observed after controlling for BMI, indicating the presence of

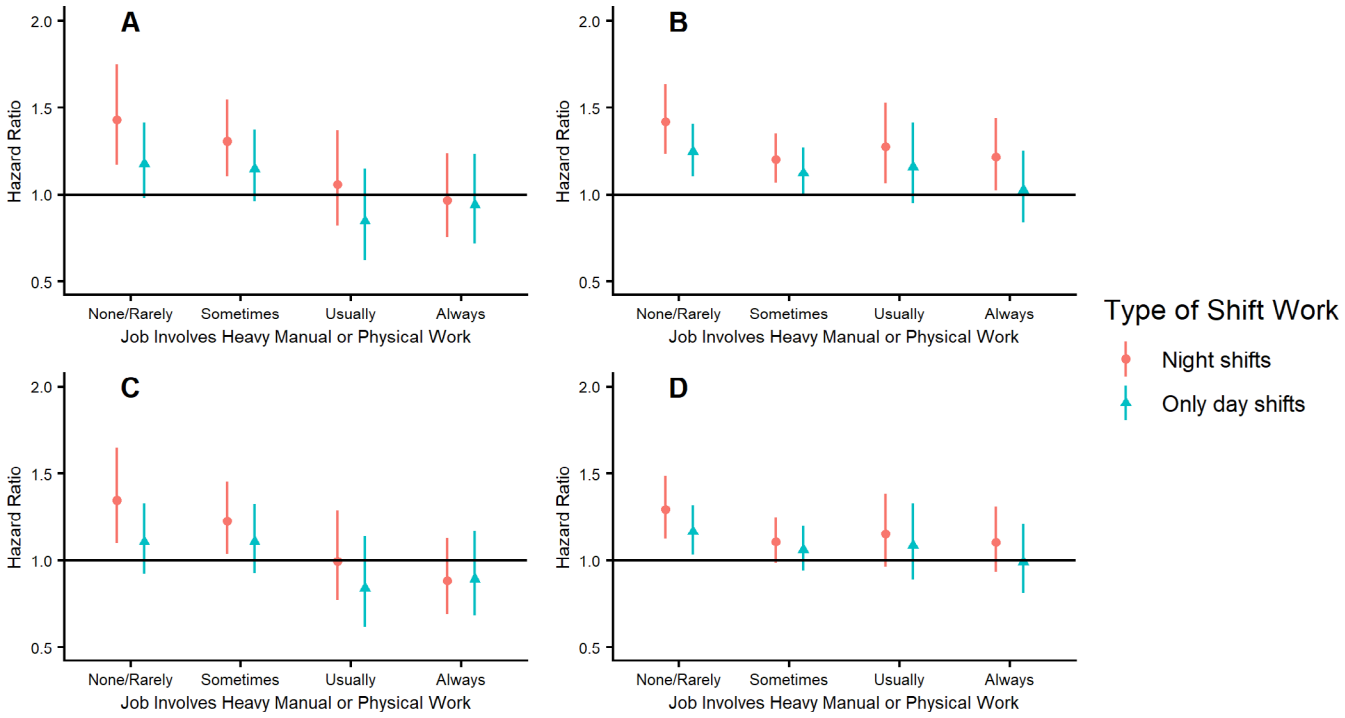

Figure 2. Shift work associations with knee osteoarthritis outcomes stratified by frequency of heavy physical work. (A) Associations with knee osteoarthritis, not adjusted for body mass index. (B) Associations with total knee arthroplasty, not adjusted for body mass index. (C) Associations with knee osteoarthritis, adjusted for body mass index. (D) Associations with total knee arthroplasty, adjusted for body mass index. All associations adjusted for age, sex, education, race, and Townsend Deprivation Index. Nonshift workers are the reference group for each hazard ratio. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.70040/abstract.

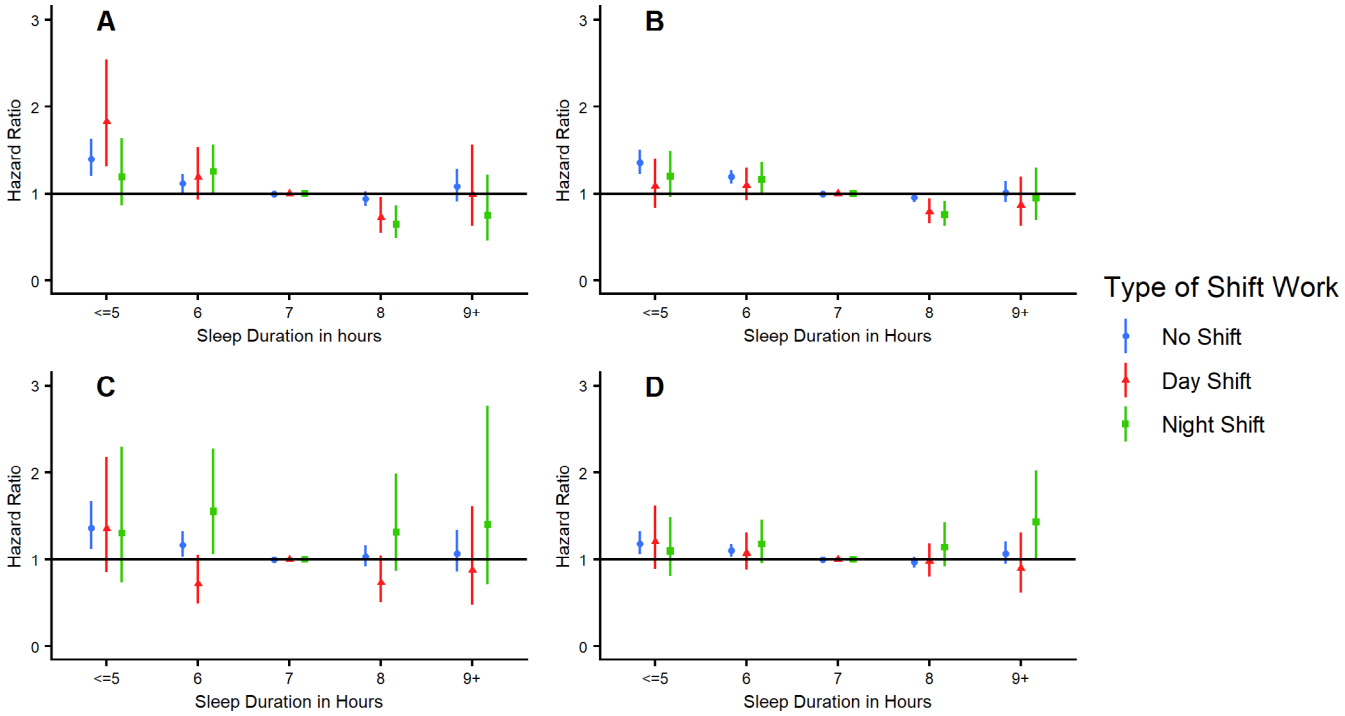

Figure 3. Sleep duration associations with osteoarthritis outcomes stratified by shift work type. (A) Associations with knee osteoarthritis. (B) Associations with total knee arthroplasty. (C) Associations with hip osteoarthritis. (D) Associations with total hip arthroplasty. All associations adjusted for age, sex, education, race, Townsend Deprivation Index, frequency of manual/physical labor, and frequency of occupational walking/standing. Participants reporting an average of 7 hours of sleep per day are the reference group for each hazard ratio.

BMI-independent mechanisms through which these factors affect OA. Short sleep and poor sleep had similar effects on knee and hip OA endpoints, whereas night shift work was only associated with knee OA, potentially indicating that circadian clock disruptions are particularly impactful for the knee. In all, these findings point to new mechanisms that may influence OA development and, correspondingly, new potential methods of OA prevention.

Both short sleep duration and more frequent sleeplessness/insomnia were strongly associated with higher risk of all hip and knee OA endpoints. Importantly, this association remained after limiting to people without chronic knee/hip pain at the time of sleep assessment. This finding is consistent with a UKB study that found that poor sleepers, based on a composite score, had a higher risk of an *ICD-9/10* OA diagnosis.[33] Genetic predisposition to insomnia and short sleep duration has been associated with hospital-diagnosed hip and knee OA,[34] and in a large case-control study, a history of sleep disorders was associated with OA, particularly for sleep apnea.[35] Less than 1% of the UKB population reported prevalent sleep apnea at their baseline assessment, but many more may have undiagnosed or unreported sleep disorders that contribute to short sleep or poor sleep quality. Sleep apnea is strongly associated with obesity[30]; however, our study shows that even after controlling for BMI at the time of sleep assessment, having difficulty getting to sleep or sleeping through the night is associated with higher OA risk. Poor sleep is known to influence pain perception,[15,36] and the associations we identified could also be influenced by heightened OA pain leading to earlier diagnosis and referral to surgical treatment.

For participants sleeping >7 hours nightly, significant OA associations were not consistently observed. However, most participants with longer sleep were within healthy recommended ranges (7–9 hours for adults aged 26–64 years and 7–8 hours for adults aged ≥65 years).[37]

Sleep length and quality can be influenced by external demands, such as work. Shift work can disrupt the circadian clock that influences sleep and many physiologic daily rhythms. We found that shift work, particularly night shift work, was strongly associated with risk of both knee OA and TKA, even after adjusting for physical work frequency. These results agree with a study of automotive workers in which night shift workers had higher knee OA risk[22] and the cross-sectional UKB study of night shift work and knee OA.[25] Our study expands these findings to demonstrate associations with incident knee OA diagnoses and TKA across a broad array of occupations. Compared with the prior UKB study, our knee OA associations were slightly stronger after excluding participants with prevalent OA at baseline, and we demonstrated that associations are markedly stronger when models are not adjusted for BMI. These BMI-unadjusted associations likely better estimate the total effect of shift work as obesity is expected to serve as a mediator rather than a confounder of the relationship between shift work and OA.

Similar associations were not observed for hip OA or THA risk. However, this is consistent with a study in mice, in which knockout of the core clock transcription factor *BMAL1* led to progressive degeneration of the articular cartilage of the knee but not of the hip.[7] Maintenance of the circadian clock may be more critical for the knee than the hip because of differences in the mechanics of the two joints and the mechanisms through which hip OA develops (ie, hip dysplasia and impingement). Additionally, obesity is more impactful on OA of the knee than the hip,[38–40] and elevated obesity is likely one mechanism through which shift work influences OA.[41,42]

Night shift work was associated with higher knee OA and TKA risk regardless of chronotype. As night shifts were defined as shifts that included work between 12:00 AM and 6:00 AM, these schedules may have been disruptive even to workers with an "eveningness" chronotype. We did not have a "social jet lag" measure (ie, sleep and wake time differences between workdays and days off),[43] and shift work may be most influential when it causes discordances in sleep-wake cycles across days.[44] Night shift work was most strongly associated with knee OA risk among workers doing jobs that never or rarely required heavy physical work. This pattern was observed when evaluating both the relative and absolute increases in risk. Night shift associations with OA could have been attenuated for workers doing more frequent physical work if the work is less demanding during night shifts compared with day shifts.

The magnitude of the associations we identified is notable within the context of studies of other lifestyle risk factors for OA. For instance, the highest quintile of Mediterranean diet adherence compared with the lowest quintile is associated with a 10% reduction in risk of symptomatic OA, whereas the highest tertile of Dietary Approaches to Stop Hypertension diet adherence compared with the lowest tertile is associated with a 22% reduction in OA risk.[45,46] Recreational running and low-to-moderate levels of strenuous physical activity are associated with 14% to 46% reductions in OA risk.[47,48] By comparison, our results indicate that 7 hours of nightly sleep compared with <6 hours is associated with a 29% reduction in knee OA risk and a 28% reduction in hip OA risk. Similarly, jobs not requiring shift work compared with night shift work are associated with a 19% reduction in knee OA risk in our study. Thus, although sleep has not previously been considered a key lifestyle factor to address to prevent OA, improved sleep and reduction of circadian clock disruptions could be as impactful as improved diet and increased physical activity. Importantly, sleep, diet, and physical activity can all influence each other,[4] and the largest benefits may come from addressing all aspects of one's lifestyle.

There are several limitations to our study. First, the UKB population has limited racial and ethnic diversity and shows evidence of healthy volunteer bias.[49] However, representativeness is not required for internal validity of risk factor–disease associations.[50]

Nonetheless, further studies to confirm the generalizability of these findings to other populations will be important. We lacked information on symptom severity and radiographic OA measures. Consequently, we cannot evaluate whether associations with OA endpoints are driven by an exacerbation of pain symptoms, an increase in cartilage degeneration, or both. Future studies in OA-focused cohorts that collect patient-reported outcomes and imaging data may be able to provide more detailed mechanistic insights. Finally, most factors were only measured at the time of the UKB baseline assessment visit. We did not evaluate duration of short sleep, sleeplessness, or shift work, but duration may be a critical influence on the magnitude of effect. Furthermore, we could not adjust for the duration of confounders, such as manual work or obesity and cannot determine if poor sleep or night shift work caused subsequent weight gain before OA diagnosis or arthroplasty.

Important strengths of this study include data on a large general population sample that covered people in a diverse array of jobs. Simultaneous assessment of sleep characteristics, occupational characteristics, and chronotype allowed us to investigate multiple aspects of sleep and circadian rhythm in relation to subsequent OA risk. Additionally, the linkage of NHS records allowed reliable ascertainment of OA-related diagnoses and procedures. Finally, the capture of other relevant variables, such as physical labor, prevalent body pain, and socioeconomic status measures, allowed us to minimize the influence of bias on our results.

Our study indicates that disruption of daily rhythms and poor sleep are novel risk factors for OA that may influence cartilage degeneration or symptom severity both through effects on obesity and through obesity-independent pathways. We identified associations between short sleep, more frequent sleeplessness/insomnia, and night shift work with greater OA risk, particularly at the knee. Associations of shift work were strongest among workers in jobs requiring minimal physical labor, indicating that the effects of circadian disruption may be strongest among people with low baseline OA risk. As circadian rhythms and sleep can be optimized through lifestyle changes and sleep hygiene interventions, these findings point to new potential ways to prevent OA development and progression.

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 Yanik 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. Allen KD, Thoma LM, Golightly YM. Epidemiology of osteoarthritis. Osteoarthritis Cartilage 2022;30(2):184–195. doi:10.1016/j.joca.2021.04.020
2. Hatzikotoulas K, Southam L, Stefansdottir L, et al; arcOGEN Consortium; ARGO Consortium; DBDS Genomic Consortium; Estonian Biobank Research Team; FinnGen; Genes & Health Research Team; HUNT All-In Pain; Million Veteran Program; Regeneron Genetics Center. Translational genomics of osteoarthritis in 1,962,069 individuals. Nature 2025;641(8065):1217–1224. doi:10.1038/s41586-025-08771-z
3. Covassin N, Singh P, Somers VK. Keeping up with the clock: circadian disruption and obesity risk. Hypertension 2016;68(5):1081–1090. doi:10.1161/HYPERTENSIONAHA.116.06588
4. Ruan W, Yuan X, Eltzschig HK. Circadian rhythm as a therapeutic target. Nat Rev Drug Discov 2021;20(4):287–307. doi:10.1038/s41573-020-00109-w
5. Rogers N, Meng QJ. Tick tock, the cartilage clock. Osteoarthritis Cartilage 2023;31(11):1425–1436. doi:10.1016/j.joca.2023.05.010
6. Dernie F, Adeyoju D. A matter of time: circadian clocks in osteoarthritis and the potential of chronotherapy. Exp Gerontol 2021;143:111163. doi:10.1016/j.exger.2020.111163
7. Dudek M, Pathiranage DRJ, Bano-Otalora B, et al. Mechanical loading and hyperosmolarity as a daily resetting cue for skeletal circadian clocks. Nat Commun 2023;14(1):7237. doi:10.1038/s41467-023-42056-1
8. Widmyer MR, Utturkar GM, Leddy HA, et al. High body mass index is associated with increased diurnal strains in the articular cartilage of the knee. Arthritis Rheum 2013;65(10):2615–2622. doi:10.1002/art.38062
9. Song X, Ma T, Hu H, et al. Chronic circadian rhythm disturbance accelerates knee cartilage degeneration in rats accompanied by the activation of the canonical Wnt/β-catenin signaling pathway. Front Pharmacol 2021;12:760988. doi:10.3389/fphar.2021.760988
10. Kc R, Li X, Forsyth CB, et al. Osteoarthritis-like pathologic changes in the knee joint induced by environmental disruption of circadian rhythms is potentiated by a high-fat diet. Sci Rep 2015;5(1):16896. doi:10.1038/srep16896
11. Bekki H, Duffy T, Okubo N, et al. Suppression of circadian clock protein cryptochrome 2 promotes osteoarthritis. Osteoarthritis Cartilage 2020;28(7):966–976. doi:10.1016/j.joca.2020.04.004
12. Gossan N, Zeef L, Hensman J, et al. The circadian clock in murine chondrocytes regulates genes controlling key aspects of cartilage homeostasis. Arthritis Rheum 2013;65(9):2334–2345. doi:10.1002/art.38035
13. Fisch KM, Gamini R, Alvarez-Garcia O, et al. Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. Osteoarthritis Cartilage 2018;26(11):1531–1538. doi:10.1016/j.joca.2018.07.012
14. Mun CJ, Tsang S, Reid MJ, et al. Effects of sleep and circadian rest-activity rhythms on daily pain severity in women with temporomandibular disorders. Pain 2025;166(7):1487–1496. doi:10.1097/j.pain.0000000000003578
15. Schrimpf M, Liegl G, Boeckle M, et al. The effect of sleep deprivation on pain perception in healthy subjects: a meta-analysis. Sleep Med 2015;16(11):1313–1320. doi:10.1016/j.sleep.2015.07.022
16. Al Khatib HK, Hall WL, Creedon A, et al. Sleep extension is a feasible lifestyle intervention in free-living adults who are habitually short sleepers: a potential strategy for decreasing intake of free sugars? A randomized controlled pilot study. Am J Clin Nutr 2018;107(1):43–53. doi:10.1093/ajcn/nqx030
17. Tasali E, Wroblewski K, Kahn E, et al. Effect of sleep extension on objectively assessed energy intake among adults with overweight in

real-life settings: a randomized clinical trial. JAMA Intern Med 2022; 182(4):365–374. doi:10.1001/jamainternmed.2021.8098

18. Cyr M, Artenie DZ, Al Bikaii A, et al. The effect of evening light on circadian-related outcomes: a systematic review. Sleep Med Rev 2022;64:101660. doi:10.1016/j.smrv.2022.101660
19. Lowden A, Öztürk G, Reynolds A, et al. Working Time Society consensus statements: evidence based interventions using light to improve circadian adaptation to working hours. Ind Health 2019; 57(2):213–227. doi:10.2486/indhealth.SW-9
20. Bartosiak K, Schwabe M, Lucey B, et al. Sleep disturbances and disorders in patients with knee osteoarthritis and total knee arthroplasty. J Bone Joint Surg Am 2022;104(21):1946–1955. doi:10.2106/JBJS.21.01448
21. Pickering ME, Chapurlat R, Kocher L, et al. Sleep disturbances and osteoarthritis. Pain Pract 2015;16(2):237–244. doi:10.1111/papr.12271
22. Zhou M, Yang S, Guo Y, et al. Shift work and the risk of knee osteoarthritis among Chinese workers: a retrospective cohort study. Scand J Work Environ Health 2020;46(2):152–160. doi:10.5271/sjweh.3861
23. Akimova ET, Taiji R, Ding X, et al. Gene-x-environment analysis supports protective effects of eveningness chronotype on self-reported and actigraphy-derived sleep duration among those who always work night shifts in the UK Biobank. Sleep 2023;46(5):zsad023. doi:10.1093/sleep/zsad023
24. Juda M, Vetter C, Roenneberg T. Chronotype modulates sleep duration, sleep quality, and social jet lag in shift-workers. J Biol Rhythms 2013;28(2):141–151. doi:10.1177/0748730412475042
25. Hashmi A, Scott S, Jung M, et al. Associations between work characteristics and osteoarthritis: a cross-sectional study of 285,947 UK Biobank participants. Osteoarthr Cartil Open 2025;7(1):100565. doi:10.1016/j.ocarto.2025.100565
26. Lane JM, Vlasac I, Anderson SG, et al. Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. Nat Commun 2016;7(1):10889. doi:10.1038/ncomms10889
27. Taillard J, Philip P, Chastang JF, et al. Validation of Horne and Ostberg morningness-eveningness questionnaire in a middle-aged population of French workers. J Biol Rhythms 2004;19(1):76–86. doi:10.1177/0748730403259849
28. Ferguson RJ, Prieto-Alhambra D, Walker C, et al. Validation of hip osteoarthritis diagnosis recording in the UK Clinical Practice Research Datalink. Pharmacoepidemiol Drug Saf 2019;28(2):187–193. doi:10.1002/pds.4673
29. Yau MS, Dubreuil M, Li S, et al. Validation of knee osteoarthritis case identification algorithms in a large electronic health record database. Osteoarthr Cartil Open 2022;4(1):100229. doi:10.1016/j.ocarto.2021.100229
30. Tuomilehto H, Seppa J, Uusitupa M. Obesity and obstructive sleep apnea–clinical significance of weight loss. Sleep Med Rev 2013; 17(5):321–329. doi:10.1016/j.smrv.2012.08.002
31. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. 3rd ed. Lippincott Williams and Wilkins; 2008.
32. Ogilvie RP, Patel SR. The epidemiology of sleep and obesity. Sleep Health 2017;3(5):383–388. doi:10.1016/j.sleh.2017.07.013
33. Sha T, Wang Y, Zhang Y, et al. Sleep patterns, genetic susceptibility, and osteoarthritis risk: insights from the UK Biobank and external validation in the Xiangya Osteoarthritis Study. Int J Surg 2025;111(7): 4443–4453. doi:10.1097/JS9.0000000000002491
34. Ni J, Zhou W, Cen H, et al. Evidence for causal effects of sleep disturbances on risk for osteoarthritis: a univariable and multivariable Mendelian randomization study. Osteoarthritis Cartilage 2022;30(3): 443–450. doi:10.1016/j.joca.2021.11.021
35. Jacob L, Smith L, Konrad M, et al. Association between sleep disorders and osteoarthritis: a case-control study of 351,932 adults in the UK. J Sleep Res 2021;30(6):e13367. doi:10.1111/jsr.13367
36. Williams JC, Hum RM, Alam U, et al. Insomnia and short sleep duration, but not chronotype, is associated with chronic widespread pain: Mendelian randomization study. Rheumatol Int 2024;44(12):2961–2966. doi:10.1007/s00296-024-05636-y
37. Hirshkowitz M, Whiton K, Albert SM, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015;1(1):40–43. doi:10.1016/j.sleh.2014.12.010
38. Reyes C, Leyland KM, Peat G, et al. Association between overweight and obesity and risk of clinically diagnosed knee, hip, and hand osteoarthritis: a population-based cohort study. Arthritis Rheumatol 2016; 68(8):1869–1875. doi:10.1002/art.39707
39. Grotle M, Hagen KB, Natvig B, et al. Obesity and osteoarthritis in knee, hip and/or hand: an epidemiological study in the general population with 10 years follow-up. BMC Musculoskelet Disord 2008;9: 132. doi:10.1186/1471-2474-9-132
40. Cho HJ, Morey V, Kang JY, et al. Prevalence and risk factors of spine, shoulder, hand, hip, and knee osteoarthritis in community-dwelling Koreans older than age 65 years. Clin Orthop Relat Res 2015; 473(10):3307–3314. doi:10.1007/s11999-015-4450-3
41. Liu Q, Shi J, Duan P, et al. Is shift work associated with a higher risk of overweight or obesity? A systematic review of observational studies with meta-analysis. Int J Epidemiol 2018;47(6):1956–1971. doi:10.1093/ije/dyy079
42. Sun M, Feng W, Wang F, et al. Meta-analysis on shift work and risks of specific obesity types. Obes Rev 2018;19(1):28–40. doi:10.1111/obr.12621
43. Wittmann M, Dinich J, Merrow M, et al. Social jetlag: misalignment of biological and social time. Chronobiol Int 2006;23(1-2):497–509. doi:10.1080/07420520500545979
44. Arendt J. Shift work: coping with the biological clock. Occup Med (Lond) 2010;60(1):10–20. doi:10.1093/occmed/kqp162
45. Zhang Y, Francis EC, Xia T, et al. Adherence to DASH dietary pattern is inversely associated with osteoarthritis in Americans. Int J Food Sci Nutr 2020;71(6):750–756. doi:10.1080/09637486.2020.1722075
46. Veronese N, Koyanagi A, Stubbs B, et al. Mediterranean diet and knee osteoarthritis outcomes: a longitudinal cohort study. Clin Nutr 2019;38(6):2735–2739. doi:10.1016/j.clnu.2018.11.032
47. Alentorn-Geli E, Samuelsson K, Musahl V, et al. The association of recreational and competitive running with hip and knee osteoarthritis: a systematic review and meta-analysis. J Orthop Sports Phys Ther 2017;47(6):373–390. doi:10.2519/jospt.2017.7137
48. Chang AH, Lee JJ, Chmiel JS, et al. Association of long-term strenuous physical activity and extensive sitting with incident radiographic knee osteoarthritis. JAMA Netw Open 2020;3(5):e204049. doi:10.1001/jamanetworkopen.2020.4049
49. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol 2017; 186(9):1026–1034. doi:10.1093/aje/kwx246
50. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013;42(4):1012–1014. doi:10.1093/ije/dys223

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1220–1228
DOI 10.1002/acr.80017
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

American College of Rheumatology
Empowering Rheumatology Professionals

Does Long-Term Lower Extremity Strength Training in Adults With Knee Osteoarthritis and Varus Alignment Reduce Knee Joint Loading During Gait?

Stephen P. Messier, PhD,[1,2] Matthew N. Vigliotti, MS,[1] Paige E. Rice, PhD,[1] Brian Pietrosimone, PhD,[3] Shannon L. Mihalko, PhD,[1] Edward H. Ip, PhD,[2] Richard F. Loeser, MD,[4] David J. Hunter, MBBS, PhD,[5] Ali Guermazi, MD, PhD,[6] Ryan Hill, MS,[1] Santiago Saldana, MS,[2] Kim L. Bennell, PhD,[7] and Paul DeVita, PhD[8]

Objective. We examined whether 18 months of strength training in individuals with knee varus alignment and medial tibiofemoral osteoarthritis (OA) reduced knee joint loads during walking compared to an attention control group.

Methods. This study was a secondary analysis of a randomized clinical trial that compared the effects of strength training to a control group in adults with knee OA. For this analysis, control participants had knee varus malalignment (≥2° varus; N = 49); participants in the strength training group met the varus malalignment criterion and increased their hip abductor strength by ≥20% from baseline to 18-month follow-up (N = 39). Linear regressions were used to compare means between groups at 18 months.

Results. The strength training group had greater increases in strength in the quadriceps (45%), hamstrings (68%), and hip abductors (42%) than the control group (16%, 11%, 4%, respectively; $P < 0.05$). There were no significant differences in the mean peak internal knee abduction moment or mean peak knee compressive force between groups at 18-month follow-up. The adjusted means at 18-month follow-up for the internal knee extension moment were significantly less (27%) in the strength training group ($P_{adjusted} = 0.03$).

Conclusion. Among older adults with knee OA and varus alignment, long-term lower extremity strength training results in significant increases in strength but does not significantly alter most measures of knee joint loading during walking relative to an attention control group. The results cast doubt on whether clinically meaningful improvement in lower extremity muscle strength translates to clinically important attenuation in knee joint loading.

INTRODUCTION

Arthritis is the leading cause of disability among adults, accounting for $304 billion in annual lost earnings and medical care costs.[1] Osteoarthritis (OA) is the most common type of arthritis,[2] and the knee joint accounts for 85% of the total burden of the disease.[3] OA is a progressive disease that affects the entire joint and surrounding tissues including cartilage, bone, synovial membrane, ligaments, muscles, and tendons, resulting in pain, stiffness, inflammation, and limited mobility. Symptoms related to OA have a negative effect on an individual's daily function, overall health, and quality of life.[4–6] Identifying strategies that can reduce disease incidence and limit disease progression are important public health objectives.[7]

The relationship between OA and muscle strength may be bidirectional; OA has a negative impact on muscle strength, but persistent muscle weakness may also hasten OA progression.[8] Compared to healthy controls, hip and knee isometric and concentric strength are significantly attenuated in patients with knee OA, generally between 16% to 27% less.[9,10] Eccentric strength deficits compared to controls can range as high as 76%.[10] Meta-analysis data also indicate that low knee extensor and flexor strength worsen structural knee OA, although the authors noted that the included studies had low certainty of evidence.[8] Possible

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant 1R01-AR-059105-01), and National Institute of Aging, NIH (grant P30-AG-21332) .

[1]Wake Forest University, Winston-Salem, North Carolina; [2]Wake Forest School of Medicine, Winston-Salem, North Carolina; [3]University of North Carolina, Chapel Hill; [4]University of North Carolina School of Medicine, Chapel Hill; [5]University of Sydney, Sydney, New South Wales, Australia; [6]Boston University School of Medicine, Boston, Massachusetts; [7]The University of Melbourne, Parkville, Victoria, Australia; [8]East Carolina University, Greenville, North Carolina.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80017.

Address correspondence via email to Stephen P. Messier, PhD, at messier@wfu.edu.

Submitted for publication March 4, 2025; accepted in revised form January 30, 2026.

SIGNIFICANCE AND INNOVATION

- Increased hip abductor strength significantly reduced contralateral pelvic drop compared to the control group; however, the combination of reduced pelvic drop and increased hip strength did not translate to a reduction in the peak internal knee abduction moment.
- Strength training reduces the peak internal knee extension moment, a kinetic variable that impacts the total knee compressive force during walking.
- Among older adults with knee osteoarthritis and varus alignment, long-term lower extremity strength training results in significant increases in strength compared to an attention control group but does not significantly alter most measures of knee joint loading during walking. The results cast doubt on whether clinically meaningful improvement in lower extremity muscle strength translates to clinically important attenuation in knee joint loading.

mechanisms that may explain the relationship between muscle weakness and knee OA include altered joint loading and the lack of muscle strength necessary to absorb forces that are transferred across the joint, resulting in increased coactivation and a disruption to cartilage/bone homeostasis.[11] Conversely, low strength may result in joint underloading or load redistribution, which can also be detrimental for cartilage/bone health.[12,13]

Frontal plane knee alignment plays a role in knee joint load distribution, with knee varus malalignment resulting in an increase in medial compartment joint loading.[14,15] People with knee OA and varus-aligned knees have greater knee joint loads approximated by the internal knee abduction moment.[16] Varus-aligned knees result in a 33% increase in the loads on the medial compartment compared to individuals with neutral-aligned knees. These larger loads on the medial portion of the knee joint may cause cartilage, bone, and other synovial joint alterations.[17] Varus malalignment is also related to knee OA progression and degradative changes to the joint.[18,19] Sharma et al[20] found that varus alignment measured at baseline was correlated with a four-fold increased risk in knee OA progression at 18-month follow-up.

Chang et al[21] discovered a greater internal hip abduction moment at baseline was associated with decreased risk of medial tibiofemoral OA progression. For every one unit increase in the internal hip abduction moment, there was a 50% decrease in the risk of OA progression. Consequently, hip muscle strengthening to increase the hip abduction moment, reduce contralateral pelvic drop, and thus shorten the distance from the knee joint center to the projection of the ground reaction force became a popular therapeutic recommendation, although empirical evidence supporting this mechanism is unreported.[22] A 9% reduction in the internal knee abduction moment was attributed to an exercise regimen that strengthened the hip and knee musculature. Not all hip abductor strengthening interventions, however, found supporting results. Studies that specifically targeted hip abductor strengthening for individuals with knee OA and varus alignment found that those in the strengthening group increased hip muscle strength, improved physical function, and reduced knee pain but did not reduce medial knee joint loads.[23,24] Similarly, improved clinical outcomes without a change in joint mechanics was seen in a short-term quadriceps strengthening clinical trial.[25] Taken together, these studies suggest that improved joint mechanics may not be essential for muscle strengthening to have a positive effect on clinical outcomes in people with knee OA. The short duration of these interventions (two to three months), however, was an important limiting factor.

We hypothesized that individuals with knee varus malalignment and medial tibiofemoral OA with at least a 20% improvement in hip abductor muscle strength consequent to 18 months of strength training will have attenuated knee joint loading during walking, including lower internal knee abduction and knee extension moments, and reduced peak bone-on-bone knee joint compressive forces, compared to an attention control group.

METHODS

Study design. START was an assessor-blinded, single-center (Wake Forest University, Winston-Salem, North Carolina), randomized clinical trial with three parallel groups followed for 18 months. The trial protocol is published elsewhere.[26] The Human Subjects Committee of Wake Forest Health Sciences approved this protocol. Participants provided written informed consent.

Participants. Participants in this randomized clinical trial (N = 377) were ambulatory, community-dwelling adults, age ≥50 years with self-reported disability due to knee OA. Enrollment occurred between July 2012 and February 2016. Inclusion criteria for the parent trial were as follows: (1) mild to moderate radiographic medial tibiofemoral OA (Kellgren–Lawrence score = 2, 3) in at least one knee and no to moderate patellofemoral OA noted on skyline view radiographs, (2) neutral (–2° valgus to 2° varus) or varus (≤10° varus) mechanically aligned knee,[27] (3) body mass index (BMI) 20 to 45 kg/m^2, and (4) no participation in formal strength training >30 minutes/wk in the past six months. All participants continued their medications, including nonsteroidal anti-inflammatory drugs. If pain decreased, they could reduce medications with their physician's approval. Exclusion criteria included the following: (1) severe patellofemoral OA (joint space width (JSW) = 3 using the Osteoarthritis Research Society International atlas from skyline x-ray views[28]), (2) lateral > medial JSW, and (3) <20 on the Montreal Cognitive Assessment.[29]

This secondary analysis was limited to START participants who (1) were randomized to the control group with ≥2° varus alignment, (2) were randomized to either strength training group with ≥2° varus alignment and increased their hip abductor strength by at least 20% over the 18-month training period, and (3) had all available strength and biomechanical gait data at baseline and 18-month follow-up (N = 88). The 20% strength improvement criterion was to ensure a clear and clinically relevant distinction in strength improvement between the intervention and control groups.

Interventions. *Strength training.* Group strength training sessions were conducted three times weekly over 18 months. Sessions consisted of a 5-minute warm-up, 40-minute training, and 15-minute cool-down supervised by staff interventionists. The program included six lower body exercises, with each leg exercised separately to prevent unequal load distribution between the limbs, and four upper body and core exercises.

The high-intensity group performed three sets of each exercise beginning at 75% of one repetition maximum (1RM) with eight repetitions/set for two weeks, progressing to 80% of the 1RM with eight repetitions/set for weeks 3 to 4, 85% of the 1RM with six repetitions/set for weeks 5 to 6, and 90% of the 1RM with four repetitions/set for weeks 7 to 8. Week 9 was a taper week with alternate exercises and establishing new 1RMs for each exercise. This nine-week block was repeated using the new 1RM values. The low-intensity group used the same nine-week block pattern but performed three sets of 15 repetitions at 30% to 40% 1RM of the exercises described above. The target workload/total volume performed during these nine-week cycles was the same regardless of whether the participant was assigned to the high-intensity or low-intensity group.[26]

Attention control group. Participants attended 60-minute group workshops biweekly for the first six months and monthly thereafter (total of 24 sessions over 18 months). Details of the control intervention are included in the study design publication.[26]

Measurements and procedures. Bilateral posteroanterior (semiflexed view) weight-bearing knee x-rays were acquired to determine the diagnosis of OA. The Kellgren–Lawrence grading scale was used to determine OA diagnosis/severity (0 to 4 scale, with 0 = no OA and 4 = severe OA).[30] A full-length lower extremity coronal view x-ray was used to determine frontal plane knee alignment.[27] The measurement for knee alignment was determined by lines drawn from the femoral head to the knee and from the center of the ankle surface to the knee. Alignment was characterized as neutral (angle = 0°), varus (angle >0° in the direction of a bow-legged appearance), or valgus (angle <0° in the direction of a knock-knee appearance).

A six-camera motion capture system was configured to collect three-dimensional kinematic data at 60 Hz. Marker trajectories were captured, tracked, and processed in Cortex 3.0, in which the raw data were smoothed using a fourth-order low-pass Butterworth filter with a 6-Hz cutoff to reduce high-frequency noise while preserving physiologically meaningful motion. Ground reaction forces were recorded using two six-channel force platforms integrated with the motion capture system, allowing synchronized kinetic data acquisition at 1,000 Hz. These analog force signals were low-pass filtered with a fourth-order Butterworth filter (cutoff 20 Hz).

The filtered motion capture and force-plate data were then imported into Visual3D (C-Motion) for biomechanical analysis. A rigid-body, segmental model was constructed based on a modified Cleveland Clinic marker set, with each segment defined by its own local coordinate system and inertial parameters calculated using Dempster-based regression equations and geometric assumptions. Pelvic and hip joint centers were estimated using the Bell and Brand regression approach, and marker placement on medial and lateral anatomic landmarks facilitated the definition of anatomic reference frames.[31,32] Joint angles were computed using a Cardan rotation sequence (X–Y–Z), corresponding to flexion/extension, abduction/adduction, and internal/external rotation. Using the synchronized three-dimensional marker trajectories and ground reaction forces, net internal joint forces and moments were calculated using a routine Newton–Euler inverse-dynamics approach in Virtual3D. Participants were instructed to walk at their normal pace, as if walking purposefully around town. Six practice trials were conducted at this self-selected walking speed, and the average time was calculated. Speed was recorded and monitored using a wireless infrared timing gate (Dual Beam System TF100, TracTronix).[26]

Kinetic outcomes included comparison of hip and knee joint loading, including peak internal hip abduction moment, peak internal knee abduction moment, and peak internal knee extension moment (note: the internal moments are presented in this analysis). Kinematic outcomes of interest included maximum frontal plane contralateral (side with the less affected knee) pelvic drop during the first 20% of the stance phase and sagittal plane knee angle.

Tibiofemoral knee compressive force was calculated using a reduction approach with a statically determinant musculoskeletal model.[33] The input parameters included the linear position of the lower extremity and the internal joint moments and reactions forces. The original sagittal plane model was improved by including a frontal plane component, which was fully described previously.[34] The model's muscle and knee joint force predictions compared favorably in terms of maximum forces and the shape of the force by time curves with those from an electromyography-based model and an optimization-based model.[34] Further, the predicted first and second peak tibiofemoral forces during walking in a person with a prosthetic knee were within 7% and 3% of those measured within a knee prosthesis.[26] Finally, Dumas et al[35] independently applied our model to the gait data from five people with knee prostheses and compared the muscle and knee

joint force predictions from our model with those measured in the prostheses during walking. They concluded that our model's predictions matched the measured data in both curve shape and force magnitude, that errors from our predictions matched those from optimization-based models, and that our model is appropriate for osteoarthritic populations.

Several measures were used to assess baseline clinical characteristics of the two groups. The Likert version of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[36] was used to assess knee pain. It assesses knee pain over the last 48 hours.[37] The total score ranges from 0 to 20 (higher scores indicate greater pain). Pain categories on a 0 to 20 scale are 2 to 8 for mild, >8 to 14 for moderate, and >14 to 20 for severe (transformed from the 0–10 scale reported by Kapstad et al[38]). WOMAC physical function assessed the degree of difficulty with activities of daily living in the last 48 hours with the total score range 0 to 68; higher scores indicate poorer function, with a score ≥21 indicating physical work limitations.[39] Six-minute walk distance, a clinical measure of mobility, assessed the distance a participant could walk along a standardized walkway in six minutes.[40]

Bilateral isokinetic strength of the knee extensors and flexors was assessed using a HUMAC NORM dynamometer (Computer Sports Medicine, Inc) with participants seated and the knee positioned at 90° flexion. Participants performed maximal concentric contractions from 90° to 30° and back at an angular velocity of 30°/s. Bilateral hip abductor strength was evaluated in a side-lying position with a hip abduction attachment secured to the femur. Participants performed maximal abductions from 0° to 30° at the same angular velocity (30°/s). For both assessments, five trials were performed with 60 seconds of rest between each trial. The top three trials were used for analysis. Additional trials were collected if the variability among the top three trials exceeded 10%.

Statistical analyses. Primary analyses were conducted using complete-case data. Of the 377 START participants, 107 were eligible based on the study's inclusion criteria—of these, 88 (49 in the control group, 39 in the strength training group) had all available baseline and 18-month strength and biomechanical data. Participants randomized to the high- and low-intensity strength training protocols that met the varus alignment, and 18-month hip abductor strength improvement inclusion criteria were combined into one strength training group for this secondary analysis.

Descriptive measures were used to summarize the baseline demographic and clinical characteristics. These characteristics included age, BMI, weight, height, sex, ethnicity, race, baseline Kellgren–Lawrence score, patellofemoral OA, WOMAC pain, WOMAC function, usual gait speed, and the six-minute walk distance. Linear regressions with adjustments for baseline outcome, age, sex, gait speed, and 18-month body weight were used to compare quadriceps, hamstrings, and hip abductor strength; internal knee abduction, hip abduction, and knee extension moments; and knee compressive force between the control and strength training groups at 18-month follow-up. The *P* value was set at ≤0.05 for statistical significance. All statistical analyses were performed using SAS version 9.4 (SAS Institute).

RESULTS

Table 1 shows the characteristics of the participants by group. Most participants were White and male with obesity and mild to moderate knee OA.

Strength. Unadjusted hip abductor strength increased by 2.9 ± 2.0 Nm (4% ± 3%) and by 24.6 ± 2.2 Nm (42% ± 8%) in the control and strength training groups, respectively. Knee extensor strength increased by 11.8 ± 2.8 Nm (16% ± 5%) in the control group and 30.8 ± 4.7 Nm (45% ± 8%) in the strength training group; knee flexor strength increased by 4.2 ± 1.9 Nm (11% ± 4%) in the control group and 22.9 ± 3.4 Nm (68% ± 7%) in the strength training group.

Table 1. Baseline descriptive characteristics of the control and strength training groups*

Variable	Control (n = 49), mean ± SD or N (%)	Strength training (n = 39), mean ± SD or N (%)
Age, y	64.2 ± 6.3	62.6 ± 6.3
BMI, kg/m^2	31.6 ± 5.3	31.7 ± 4.3
Body mass, kg	88.5 ± 15.3	94.0 ± 20.0
Height, cm	167.4 ± 9.4	171.6 ± 11.8
Female	18 (36.7)	18 (46.2)
Hispanic	0 (0.0)	1 (2.8)
Race		
Black	9 (18.4)	7 (17.9)
White	39 (79.6)	31 (79.5)
Other	1 (2.0)	1 (2.6)
KL grade 2	16 (32.7)	17 (43.6)
KL grade 3	28 (57.1)	20 (51.3)
KL grade 4	5 (10.2)	2 (5.1)
PF osteoarthritis	44 (89.8)	34 (87.2)
Knee varus alignment, degrees	4.3 ± 1.8	4.1 ± 1.8
WOMAC pain (0–20)	6.9 ± 2.4	7.1 ± 2.9
WOMAC function (0–68)	24.5 ± 9.5	26.0 ± 11.1
Usual gait speed, m/s	1.26 ± 0.15	1.25 ± 0.14
Six-minute walk, m	488.0 ± 104.9	483.9 ± 72.1

* The KL scale ranges from 0 to 4. A grade of 2 or greater indicates definite osteoarthritis. PF osteoarthritis was measured from skyline view radiograph using the Osteoarthritis Research Society International scale (0 = none, 1 = mild, 2 = moderate, and 3 = severe). The WOMAC includes 5 questions on pain and 17 questions on physical function (scale for each question ranges from 0 = no symptoms to 4 = extreme symptoms). Composite scores for pain range from 0 to 20 and for function 0 to 68. BMI, body mass index; KL, Kellgren–Lawrence; PF, patellofemoral; WOMAC, Western Ontario and MacMasters Universities Osteoarthritis Index.

The adjusted mean value for hip abductor strength at 18 months was 70.3 ± 1.7 Nm for the control group and 85.8 ± 1.8 Nm for the strength training group, a difference of 22% that was statistically significant (mean difference 15.6 Nm, 95% confidence interval [CI] 10.6–20.5, $P \leq 0.0001$). There was a 13% difference in knee extensor strength between groups at 18 months (mean difference 11.3 Nm, 95% CI 3.2–19.4, $P = 0.007$) in favor of the strength training group. The adjusted difference in knee flexor strength between groups at 18 months differed by 26% and favored the strength training group (mean difference 11.7 Nm, 95% CI 6.5–17.0 Nm, $P \leq 0.0001$; Table 2).

Gait biomechanics. The sample means for the unadjusted internal hip abduction moments increased in both the control (1.5 ± 3.6 Nm, 2% ± 2%) and strength training (11.3 ± 3.8 Nm, 15% ± 6%) groups with no significant difference between the groups in adjusted 18-month means (mean difference 7.0 Nm, 95% CI –2.2 to 16.2, $P = 0.13$). The mean increase in the unadjusted peak internal knee abduction moment across the 18-month intervention period was 0.7 ± 1.6 Nm (1.7% ± 2%) for the control group and 2.2 ± 1.9 Nm (5% ± 4%) for the strength training group. Adjusted peak internal knee abduction moment values at follow-up were not significantly different between the groups (mean difference 0.1 Nm, 95% CI –4.2 to 4.3, $P = 0.98$; Table 3; Figure 1).

The adjusted means at 18-month follow-up for the peak internal knee extension moment were significantly lower in the strength training group compared to the control group (mean difference –9.5 Nm, 95% CI –18.3 to –0.7, $P = 0.03$; Table 3). The adjusted mean values for the peak knee compressive forces were 2,547 ± 94 N (2.95 body weights) for the control group and 2,415 ± 103 N (2.65 body weights) for the strength training group and were statistically similar between groups (mean difference –133 N, 95% CI –407 to 142, $P = 0.34$; Figure 2).

Contralateral pelvic drop (the side with the least affected knee) at 18-month follow-up, adjusted for baseline gait speed, age, gender, and baseline differences in the outcome, was significantly less in the strength training group relative to the control group (mean difference 0.7 degrees, 95% CI –1.45 to –0.01, $P = 0.047$; Table 3). Within the strength training group, the mean peak knee flexion angles during the first 20% of stance were similar at baseline (11.7 ± 5.6 degrees) and 18 months (12.2 ± 6.4 degrees; Figure 3A). Figure 3B shows a within group reduction in the unadjusted internal knee extension moment during the first 20% of stance in the strength training group.

DISCUSSION

Among older adults with knee OA and varus alignment, 18 months of strength training that leads to a significant improvement in hip and knee strength does not have a significant effect on the magnitude of the peak hip and knee abduction moments, the

Table 2. Baseline and 18-month follow-up mean muscle strength for the control and strength training groups adjusted for covariates*

Variable	Control				Strength training				Between groups	
	Baseline (n = 49), mean (±SD)	18 months (n = 49), mean (±SD)	% Δ, mean (±SE)	18 months adj, mean (±SE)	Baseline (n = 39), mean (±SD)	18 months (n = 39), mean (±SD)	% Δ, mean (±SE)	18 months adj, mean (±SE)	Mean difference (95% CI)	P^a
Hip abductor strength, Nm	66.2 (±23.6)	69.1 (±20.9)	4 (±3)	70.3 (±1.7)	58.6 (±22.8)	83.2 (±28.9)	42 (±8)	85.8 (±1.8)	15.6 (10.6–20.5)	≤0.0001
Knee extensor strength, Nm	73.9 (±34.6)	85.7 (±32.9)	16 (±5)	89.8 (±2.8)	69.2 (±28.2)	100.0 (±34.1)	45 (±8)	101.1 (±3.0)	11.3 (3.2–19.4)	0.007
Knee flexor strength, Nm	38.7 (±23.0)	42.9 (±19.2)	11 (±4)	46.2 (±1.8)	33.7 (±18.1)	56.6 (±25.6)	68 (±7)	57.9 (±1.9)	11.7 (6.5–17.0)	≤0.0001

* % Δ (percent change) = (18-month unadjusted mean – baseline mean) ÷ baseline mean × 100. adj, adjusted; CI, confidence interval.
[a] *P* values signify statistical differences in the mean values between groups at 18 months adj for age, sex, body weight, and strength assessments at baseline.

Table 3. Gait outcomes for the control and strength training groups at baseline and 18-month follow-up adjusted for covariates*

Variable	Control group				Strength training group				Between groups	
	Baseline, mean (±SD)	18 months, mean (±SD)	18 months adj, mean (±SE)	% Δ, mean (±SE)	Baseline, mean (±SD)	18 months, mean (±SD)	18 months adj, mean (±SE)	% Δ, mean (±SE)	Mean difference (95% CI)	P[a]
Internal hip abduction moment, Nm	78.5 (±19.8)	80.0 (±24.0)	80.6 (±3.1)	1.9 (±2.0)	76.5 (±22.3)	87.8 (±24.1)	87.6 (±3.4)	14.8 (±6.2)	7.0 (−2.2 to 16.2)	0.13
Internal knee abduction moment, Nm	41.0 (±14.9)	41.7 (±16.6)	43.2 (±1.5)	1.7 (±1.9)	41.0 (±14.6)	43.2 (±13.5)	43.2 (±1.6)	5.4 (±3.9)	0.1 (−4.2 to 4.3)	0.98
Internal knee extension moment, Nm	41.0 (±19.2)	43.9 (±25.7)	44.8 (±3.0)	7.1 (±3.8)	42.2 (±19.9)	35.8 (±17.8)	35.3 (±3.3)	−15.2 (±6.2)	−9.5 (−18.3 to −0.7)	0.03
Knee compressive force, N	2,265 (±615)	2,476 (±843)	2,547 (±94)	9.3 (±4)	2,399 (±808)	2,451 (±805)	2,415 (±103)	2.2 (±2.6)	−133 (−407 to 142)	0.34
Pelvic drop, degrees	1.7 (±1.2)	2.1 (±1.4)	2.2 (±0.2)	24.1 (±6.4)	2.4 (±1.6)	1.6 (±2.0)	1.5 (±0.3)	−50 (±8.4)	−0.7 (−1.5 to −0.0)	0.047
Gait speed, m/s	1.3 (±0.2)	1.3 (±0.2)	1.2 (±0.0)	0.6 (±1.2)	1.3 (±0.1)	1.3 (±0.2)	1.3 (±0.0)	4.4 (±3.3)	0.0 (−0.0 to 0.1)	0.25
Weight, kg	88.5 (±15.4)	88.1 (±16.2)	90.1 (±2.3)	−0.5 (±0.9)	94.1 (±20.0)	92.8 (±19.7)	93.1 (±2.6)	−1.4 (±1.8)	2.4 (−4.4 to 9.3)	0.48

* % Δ = (18-month unadjusted mean − baseline mean) ÷ baseline mean × 100. adj, adjusted; CI, confidence interval.
[a] *P* values signify statistical differences in the mean values between groups at 18 months adj for baseline outcome, age, sex, gait speed, and 18-month body weight.

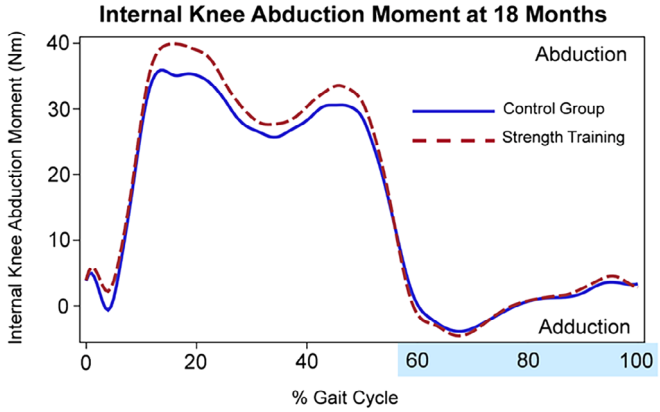

Figure 1. Mean unadjusted internal knee abduction moments at 18-month follow-up across a gait cycle (touchdown of one foot [0%] to touchdown of the same foot [100%]) for the control and strength training groups.

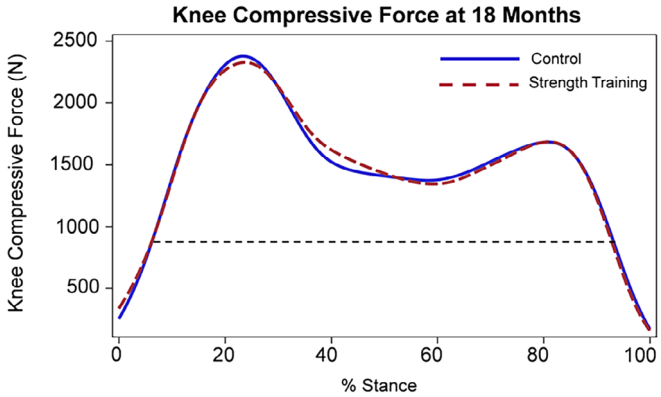

Figure 2. Mean 18-month knee compressive forces during the stance phase of gait for the control and strength training groups. The horizontal dashed line represents mean body weight.

latter being a surrogate measure of medial compartment joint loading. The hip musculature has become a popular addition to strengthening interventions for patients with OA to counterbalance external loads, reduce contralateral pelvic drop, and in turn cause a subsequent decrease in the internal knee abduction moment.[21,41] Bennell et al[23] showed that a short-term, three-month home-based hip abductor strength training intervention (13% increase in strength) reduced pain and improved function but had no effect on the internal abduction moment compared to control. Sled et al[24] found similar results from a two-month home-based hip abductor strength training intervention. The current study was longer in duration (18 months) than these studies and included only participants in the strength training group that improved hip abductor strength by at least 20% from baseline; mean improvement was 42% compared to a 4% increase in the control group.

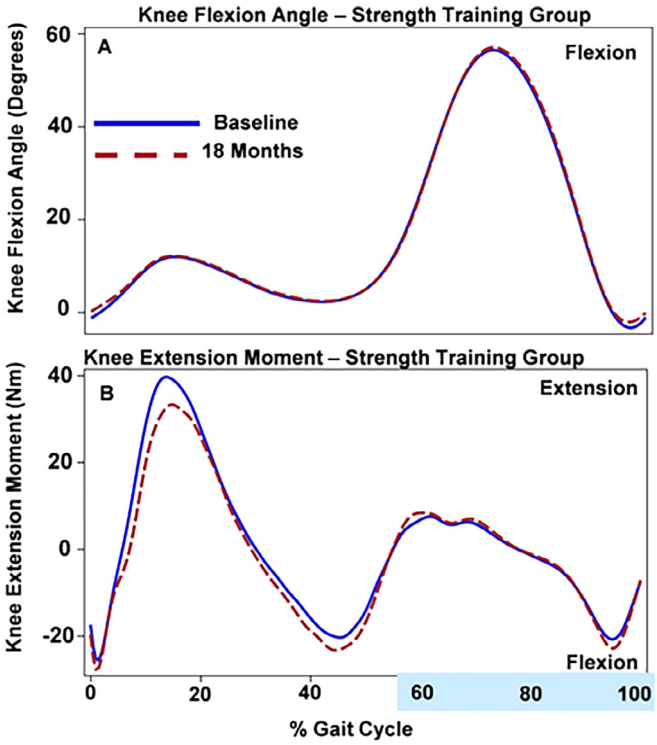

Figure 3. (Top) Mean knee flexion angle, and (bottom) mean internal knee extension moment at baseline and 18 months in the strength training group across the gait cycle (0% touchdown to 100% touchdown of the same foot; 0%–60% stance and 60%–100% swing phase).

The increased strength significantly reduced contralateral pelvic drop compared to the control group; however, the combination of reduced pelvic drop and increased hip strength did not translate to a reduction in the peak internal knee abduction moment. In contrast, Thorp et al conducted a four-week lower extremity strength training intervention exercising the knee extensors, knee flexors, and hip abductors in six participants with radiographic knee OA and found that the knee abduction moment decreased by 9%. The small sample size, short intervention duration, and lack of a comparator group limited the generalizability and the strength of evidence of the results.

An internal knee extension moment refers to the torque produced by the extensor muscles relative to that produced by the flexor muscles when the extensors contract eccentrically to control knee flexion after heel strike, essentially acting as a measure of the force being applied to the joint during this time.[42] The strength training group increased knee extensor strength by 45% compared to 16% for the control group. This resulted in a significant decrease of 27% in the adjusted internal knee extension moment during the first 20% of the gait cycle. The knee flexion angle did not change from baseline to follow-up, suggesting that the reduced knee extension moment may be reflective of a reduction in medial compartment contact force.

The benefits of load reduction on clinical outcomes are evident in long-term dietary weight-loss clinical trials[43,44] and in studies of OA disease progression.[15] In contrast, higher internal knee extension moments are characteristic of healthy control groups compared to people with knee OA.[45] These higher joint moments are likely affected by the faster walking speeds of people with healthy knees. Nevertheless, it is important to note that people with healthy knees tend to have greater joint loads compared to people with knee OA.[46]

Clinical trials of diet and exercise have shown reductions in bone-on-bone knee joint compressive forces with weight loss.[47] In our weight-stable (mean change = 1.4%) strength training group, however, there was no reduction in compressive forces. These data cast doubt as to whether strengthening the knee extensor muscles has any shock-absorbing benefit in osteoarthritic gait.

This study has several limitations. Our study samples were quite specific and therefore our results may be limited to a similar population. Namely, these results may be generalizable to individuals who are White and male with knee OA, obesity, and middle to older age, have varus malalignment, and who have made substantial increases in their lower extremity strength. We used musculoskeletal modeling to estimate knee joint compressive forces. All such models have simplifying assumptions about the tissues they describe, including bones, ligaments, and muscles, and the joint properties.[48] Despite these limiting assumptions, our model predictions have been shown to be reasonable. Healthy older adults and older adults with knee OA typically use a greater percentage of their maximum strength during gait than healthy young adults; changes in submaximal functional muscle capacity were not assessed.[49] We hypothesized that greater muscle strength would lead to lower joint loads. However, lower joint loads may represent redistributed loads, either to other parts of the same joint or to other joints. Our analysis did not distinguish between load attenuation and load redistribution. Finally, this study was a secondary analysis from a randomized clinical trial and included a subset of participants, hence randomization was no longer preserved. Despite these limitations, we propose the data and results are reasonably robust and accurately reflect responses in a real-world setting.

Among older adults with knee OA and varus alignment, long-term lower extremity strength training results in significant increases in strength but does not significantly alter most measures of knee joint loading during walking relative to an attention control group. The results cast doubt on whether clinically meaningful improvement in lower extremity muscle strength translates to clinically important attenuation in knee joint loading.

ACKNOWLEDGMENTS

We thank the START research staff and the START participants for their important contributions. The participants did not receive

compensation for their contributions, and the staff did not receive additional compensation besides their salaries.

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 Messier 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. Theis KA, Murphy LB, Guglielmo D, et al. Prevalence of arthritis and arthritis-attributable activity limitation – United States, 2016-2018. MMWR Morb Mortal Wkly Rep 2021;70(40):1401–1407. doi:10.15585/mmwr.mm7040a2
2. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21(9):1145–1153. doi:10.1016/j.joca.2013.03.018
3. Liu M, Jin F, Yao X, et al. Disease burden of osteoarthritis of the knee and hip due to a high body mass index in China and the USA: 1990-2019 findings from the global burden of disease study 2019. BMC Musculoskelet Disord 2022;23(1):63. doi:10.1186/s12891-022-05027-z
4. Abramoff B, Caldera FE. Osteoarthritis. Med Clin North Am 2020;104(2):293–311. doi:10.1016/j.mcna.2019.10.007
5. Lopes DG, Costa D, Cruz EB, et al. Association of physical activity with physical function and quality of life in people with hip and knee osteoarthritis: longitudinal analysis of a population-based cohort. Arthritis Res Ther 2023;25:14.
6. Primorac D, Molnar V, Rod E, et al. Knee osteoarthritis: a review of pathogenesis and state-of-the-art non-operative therapeutic considerations. Genes (Basel) 2020;11(8):854. doi:10.3390/genes11080854
7. Pre-Competitive Consortium for Osteoarthritis; Osteoarthritis Research Society International. OA as a serious disease. Osteoarthritis Research Society International. December 1, 2016. Accessed August 31, 2022. https://oarsi.org/research-oarsi-success/oa-serious-disease
8. Patterson BE, Girdwood MA, West TJ, et al. Muscle strength and osteoarthritis of the knee: a systematic review and meta-analysis of longitudinal studies. Skeletal Radiol 2023;52(11):2085–2097. doi:10.1007/s00256-022-04266-4
9. Hinman RS, Hunt MA, Creaby MW, et al. Hip muscle weakness in individuals with medial knee osteoarthritis. Arthritis Care Res (Hoboken) 2010;62(8):1190–1193. doi:10.1002/acr.20199
10. Alnahdi AH, Zeni JA, Snyder-Mackler L. Muscle impairments in patients with knee osteoarthritis. Sports Health 2012;4(4):284–292. doi:10.1177/1941738112445726
11. Raghava Neelapala YV, Bhagat M, Shah P. Hip muscle strengthening for knee osteoarthritis: a systematic review of literature. J Geriatr Phys Ther 2020;43(2):89–98. doi:10.1519/JPT.0000000000000214
12. Souza RB, Baum T, Wu S, et al. Effects of unloading on knee articular cartilage T1rho and T2 magnetic resonance imaging relaxation times: a case series. J Orthop Sports Phys Ther 2012;42(6):511–520. doi:10.2519/jospt.2012.3975
13. Gardinier ES, Di Stasi S, Manal K, et al. Knee contact force asymmetries in patients who failed return-to-sport readiness criteria 6 months after anterior cruciate ligament reconstruction. Am J Sports Med 2014;42(12):2917–2925. doi:10.1177/0363546514552184
14. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med 2010;26(3):355–369. doi:10.1016/j.cger.2010.03.001 Erratum in: Clin Geriatr Med 2013;29:ix.
15. Hurwitz DE, Ryals AB, Case JP, et al. The knee adduction moment during gait in subjects with knee osteoarthritis is more closely correlated with static alignment than radiographic disease severity, toe out angle and pain. J Orthop Res 2002;20(1):101–107. doi:10.1016/S0736-0266(01)00081-X
16. Messier SP, Pater M, Beavers DP, et al. Influences of alignment and obesity on knee joint loading in osteoarthritic gait. Osteoarthritis Cartilage 2014;22(7):912–917. doi:10.1016/j.joca.2014.05.013
17. Tanamas S, Hanna FS, Cicuttini FM, et al. Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review. Arthritis Rheum 2009;61(4):459–467. doi:10.1002/art.24336
18. Birmingham TB, Kramer JF, Kirkley A, et al. Association among neuromuscular and anatomic measures for patients with knee osteoarthritis. Arch Phys Med Rehabil 2001;82(8):1115–1118. doi:10.1053/apmr.2001.24306
19. Brouwer GM, van Tol AW, Bergink AP, et al. Association between valgus and varus alignment and the development and progression of radiographic osteoarthritis of the knee. Arthritis Rheum 2007;56(4):1204–1211. doi:10.1002/art.22515
20. Sharma L, Dunlop DD, Cahue S, et al. Quadriceps strength and osteoarthritis progression in malaligned and lax knees. Ann Intern Med 2003;138(8):613–619. doi:10.7326/0003-4819-138-8-200304150-00006
21. Chang A, Hayes K, Dunlop D, et al. Hip abduction moment and protection against medial tibiofemoral osteoarthritis progression. Arthritis Rheum 2005;52(11):3515–3519. doi:10.1002/art.21406
22. Thorp LE, Wimmer MA, Foucher KC, et al. The biomechanical effects of focused muscle training on medial knee loads in OA of the knee: a pilot, proof of concept study. J Musculoskelet Neuronal Interact 2010;10(2):166–173.
23. Bennell KL, Hunt MA, Wrigley TV, et al. Hip strengthening reduces symptoms but not knee load in people with medial knee osteoarthritis and varus malalignment: a randomised controlled trial. Osteoarthritis Cartilage 2010;18(5):621–628. doi:10.1016/j.joca.2010.01.010
24. Sled EA, Khoja L, Deluzio KJ, et al. Effect of a home program of hip abductor exercises on knee joint loading, strength, function, and pain in people with knee osteoarthritis: a clinical trial. Phys Ther 2010;90(6):895–904. doi:10.2522/ptj.20090294
25. DeVita P, Aaboe J, Bartholdy C, et al. Quadriceps-strengthening exercise and quadriceps and knee biomechanics during walking in knee osteoarthritis: a two-centre randomized controlled trial. Clin Biomech (Bristol) 2018;59:199–206. doi:10.1016/j.clinbiomech.2018.09.016
26. Messier SP, Mihalko SL, Beavers DP, et al. Strength Training for Arthritis Trial (START): design and rationale. BMC Musculoskelet Disord 2013;14(1):208. doi:10.1186/1471-2474-14-208
27. Sharma L, Lou C, Cahue S, et al. The mechanism of the effect of obesity in knee osteoarthritis: the mediating role of malalignment. Arthritis Rheum 2000;43(3):568–575. doi:10.1002/1529-0131(200003)43:3<568:AID-ANR13>3.0.CO;2-E
28. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15 Suppl A:A1–A56.
29. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695–699. doi:

10.1111/j.1532-5415.2005.53221.x Erratum in: J Am Geriatr Soc 2019;67:1991.

30. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16(4):494–502. doi:10.1136/ard.16.4.494
31. Bell AL, Brand RA, Pedersen DR. Prediction of hip joint centre location from external landmarks. Hum Mov Sci 1989;8(1):3–16. doi:10.1016/0167-9457(89)90020-1
32. Bell AL, Pedersen DR, Brand RA. A comparison of the accuracy of several hip center location prediction methods. J Biomech 1990;23(6):617–621. doi:10.1016/0021-9290(90)90054-7
33. DeVita P, Hortobagyi T. Functional knee brace alters predicted knee muscle and joint forces in people with ACL reconstruction during walking. J Appl Biomech 2001;17(4):297–311. doi:10.1123/jab.17.4.297
34. Messier SP, Legault C, Loeser RF, et al. Does high weight loss in older adults with knee osteoarthritis affect bone-on-bone joint loads and muscle forces during walking? Osteoarthritis Cartilage 2011;19(3):272–280. doi:10.1016/j.joca.2010.11.010
35. Dumas R, Barré A, Moissenet F, et al. Can a reduction approach predict reliable joint contact and musculo-tendon forces? J Biomech 2019;95:109329. doi:10.1016/j.jbiomech.2019.109329
36. Bellamy N, Buchanan WW, Goldsmith CH, et al. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988;1:95–108.
37. Griffiths G, Bellamy N, Kean WF, et al. A study of the time frame dependency of responses to the WOMAC osteoarthritis index. Inflammopharmacology 1993;2(1):85–87. doi:10.1007/BF02663745
38. Kapstad H, Hanestad BR, Langeland N, et al. Cutpoints for mild, moderate and severe pain in patients with osteoarthritis of the hip or knee ready for joint replacement surgery. BMC Musculoskelet Disord 2008;9(1):55. doi:10.1186/1471-2474-9-55
39. Siviero P, Limongi F, Gesmundo A, et al; EPOSA Research Group. Minimal clinically important decline in physical function over one year: EPOSA study. BMC Musculoskelet Disord 2019;20(1):227. doi:10.1186/s12891-019-2593-1
40. Naylor JM, Mills K, Buhagiar M, et al. Minimal important improvement thresholds for the six-minute walk test in a knee arthroplasty cohort: triangulation of anchor- and distribution-based methods. BMC Musculoskelet Disord 2016;17(1):390. doi:10.1186/s12891-016-1249-7
41. Perry J, Burnfield J. Gait Analysis: Normal and Pathological Function. 2nd ed. Taylor & Francis Group; 2010.
42. Manal K, Gardinier E, Buchanan TS, et al. A more informed evaluation of medial compartment loading: the combined use of the knee adduction and flexor moments. Osteoarthritis Cartilage 2015;23(7):1107–1111. doi:10.1016/j.joca.2015.02.779
43. Messier SP, Mihalko SL, Beavers DP, et al. Effect of high-intensity strength training on knee pain and knee joint compressive forces among adults with knee osteoarthritis: the START randomized clinical trial. JAMA 2021;325(7):646–657. doi:10.1001/jama.2021.0411
44. Messier SP, Legault C, Mihalko S, et al. The Intensive Diet and Exercise for Arthritis (IDEA) trial: design and rationale. BMC Musculoskelet Disord 2009;10:93.
45. Astephen JL, Deluzio KJ, Caldwell GE, et al. Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity. J Orthop Res 2008;26(3):332–341. doi:10.1002/jor.20496
46. Schmitz RJ, Harrison D, Wang H-M, et al. Sagittal-plane knee moment during gait and knee cartilage thickness. J Athl Train 2017;52(6):560–566. doi:10.4085/1062-2050-52.4.05
47. Messier SP, Mihalko SL, Legault C, et al. Effects of intensive diet and exercise on knee joint loads, inflammation, and clinical outcomes among overweight and obese adults with knee osteoarthritis: the IDEA randomized clinical trial. JAMA 2013;310(12):1263–1273. doi:10.1001/jama.2013.277669
48. Wilson DR, McWalter EJ, Johnston JD. The measurement of joint mechanics and their role in osteoarthritis genesis and progression. Rheum Dis Clin North Am 2008;34(3):605–622. doi:10.1016/j.rdc.2008.05.002
49. Tung EV, Marriott KA, Laing AC, et al. The relationship between muscle capacity utilization during gait and pain in people with symptomatic knee osteoarthritis. Gait Posture 2022;94:58–66. doi:10.1016/j.gaitpost.2022.02.021

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1229–1238
DOI 10.1002/acr.80018
© 2026 American College of Rheumatology.

American College of Rheumatology
Empowering Rheumatology Professionals

Association of Wearable Sensor–Derived Gait Measures With Cartilage Damage Over Two Years Among Adults With or at Risk of Knee Osteoarthritis

Kathryn L. Bacon, PhD,[1] David T. Felson, MD, MPH,[1] S. Reza Jafarzadeh, PhD,[1] Jeffrey M. Hausdorff, PhD,[2,3,4] Eran Gazit, MSc,[3] Ali Guermazi, MD, PhD,[1] Frank W. Roemer, MD,[1,5,6] Neil A. Segal, MD, MS,[7,8] Cora E. Lewis, MD, MSPH,[9] Michael C. Nevitt, PhD,[10] and Deepak Kumar, PT, PhD,[1,11] the Multicenter Osteoarthritis Study Investigators

Objective. Gait affects knee loading. Modifying gait could reduce load and protect against cartilage loss. Our objective is to look for modifiable gait parameters and determine their relation to worsening cartilage damage.

Methods. We studied participants from the Multicenter Osteoarthritis Study (MOST) ages 45 to 90 years with, or at risk for, knee osteoarthritis (OA). Gait assessment used inertial measurement units (APDM, Inc) on the pelvis and ankles during a 20-m walk. Knee magnetic resonance imaging (MRI) was acquired at baseline and two years later. Cartilage damage worsening was assessed using MRI Osteoarthritis Knee Scores in 14 knee subregions. We examined change (yes/no) in each subregion. We used ensemble machine learning to discriminate subregions with and without cartilage damage. Predictors tested included gait variables, radiographic OA, baseline cartilage damage, age, sex, height, weight, depressive symptoms, and race/clinic site. Data were split 70% training and 30% test sets. We identified the 10 variables that, across 100 repetitions, most frequently contributed to risk of damage. We used G-computation to evaluate causal risk differences of worsening cartilage damage for each variable.

Results. We studied 1,703 participants (mean [±SD] age 61.4 [±9.4] years, 56% female). At two years, 46% had worse cartilage damage in at least one knee subregion. Of gait variables, longer step length was associated with increased risk of damage, especially in knees with more baseline damage.

Conclusion. Longer step length was associated with worse cartilage damage over two years. Interventions to shorten step length might reduce risk of worsening cartilage damage.

INTRODUCTION

Knee osteoarthritis (OA) is a leading cause of disability in adults worldwide.[1] Clinical management of knee OA primarily addresses symptoms because there are no interventions that prevent worsening of structural features, especially cartilage damage, a signature feature of the disease. There is a need for approaches that might reduce cartilage damage, especially those that also reduce pain, the main reason persons seek care for knee OA.

Knee OA is, in part, driven by altered knee loading during daily activities, and these activities may be implicated in structural worsening.[2] Specifically, gait measures reflecting greater knee joint loading during walking (eg, peak knee adduction moment,

TheMulticenter Osteoarthritis Study (MOST) was supported by the NIH, Department of Health and Human Services (grants MOST4 U19-AG-076471; P30-AR-0702571 and AG18820 [to Dr Felson]; R21-AR-074578 [to Dr Jafarzadeh]; R03-AG-060272 [to Drs Jafarzadeh and Kumar]; AG18832; AG18947 [to Dr Lewis]; and AG19069 [to Dr Nevitt]), and conducted by MOST investigators. This study was also supported by the University of California, San Francisco; The University of Iowa; The University of Alabama at Birmingham; Boston University; and National Institute on Aging, NIH. Dr Bacon's work was supported by the Rheumatology Research Foundation Investigator Award. The NIH and Rheumatology Research Foundation were not involved in study design, collection, analysis or interpretation of data, or the decision to submit this manuscript for publication.

[1]Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts; [2]Tel Aviv University, Tel Aviv, Israel; [3]Tel Aviv Sourasky Medical Center, Tel Aviv, Israel; [4]Rush University Medical Center, Chicago, Illinois; [5]Universitätsklinikum Erlangen, Erlangen, Germany; [6]Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany; [7]University of Kansas Medical Center, Kansas City; [8]The University of Iowa, Iowa City; [9]The University of Alabama at Birmingham; [10]University of California, San Francisco; [11]Boston University, Boston, Massachusetts.

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.80018).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80018.

Address correspondence via email to Kathryn L. Bacon, PhD, at kbacon@bu.edu.

Submitted for publication April 9, 2025; accepted in revised form February 2, 2026.

SIGNIFICANCE & INNOVATIONS

- Modifying gait could reduce knee load and protect against worsening cartilage damage.
- The Multicenter Osteoarthritis Study is a large cohort of adults with or at risk of knee osteoarthritis, including gait data from inertial sensors, longitudinal data on cartilage changes, and data on many confounders.
- Over two years, 46% of participants showed worsening cartilage damage in at least one knee subregion. Of gait variables, longer step length was associated with increased risk of damage, a relationship that was stronger with more baseline damage.
- Longer step length was associated with worse cartilage damage over two years, especially in those with greater cartilage damage at baseline. Interventions to shorten step length might reduce risk of worsening cartilage damage.

knee adduction moment impulse, medial peak contact force) are reported to increase the risk of structural worsening of knee OA.[2] Although these observations have informed several interventions (eg, gait retraining, orthotics), none of these interventions are strongly recommended in knee OA guidelines.[3] Gait measures previously identified as being related to structural worsening of knee OA require sophisticated motion analysis laboratories for measurement. As such, implementing interventions that target these gait measures presents a logistical challenge. Further, prior studies have been relatively small (n = 20–200), raising risks of confounding given the complex interplay among biomechanical, clinical, and other risk factors of knee OA.

Wearable inertial sensors have been validated for assessment of gait in people with knee OA and allow for implementation in large cohorts and clinical settings.[4–6] Use of wearable sensors to characterize gait in large cohorts also offers the opportunity to use machine learning that can identify gait alterations related to the outcomes without making a priori assumptions about underlying relationships among various risk factors.[7–10] We have previously identified wearable sensor–derived gait measures related to worsening pain (ie, longer step length, lower dominant frequency reflecting greater step duration and lower cadence, and greater gait asymmetry) and those related to worsening physical function (ie, lower cadence) using data from the large Multicenter Osteoarthritis Study (MOST).[11] We are unaware of studies that have examined the relation of wearable sensor–derived gait measures with structural worsening such as cartilage damage worsening.

Our objective was to identify gait characteristics associated with worsening cartilage damage over two years in a large cohort. We first used machine learning to identify both gait and nongait parameters related to worsening damage. We then determined which gait parameters were associated with worsening cartilage damage while adjusting for potential confounders.

PARTICIPANTS AND METHODS

Study population. The MOST is an observational cohort study of risk factors related to knee OA. We used data from the 144-month (baseline) and 168-month (two-year follow-up) visits of MOST. At both visits, we included data from participants from the original MOST cohort (age 50–79 years, with or at increased risk of developing knee OA at enrollment) and a new cohort enrolled at 144-month visit (age 45–69 years, Kellgren–Lawrence grades [KLGs] ≤2, with or without knee pain). The majority of participants in this study were from the new cohort in whom the disease, if present, was mild. We included participants who had both inertial sensor gait data from a 20-m walk test at the 144-month clinic visit and MRI Osteoarthritis Knee Scores (MOAKSs)[12] at the 144- and 168-month MOST visits (Figure 1). All participants signed an institutionally approved informed consent form before any study procedures.

Gait features. A wearable inertial sensor system (OPAL, APDM, Inc) was used to collect spatiotemporal gait measures during walking over ground. Sensors were worn on the trunk and each ankle. Each sensor contains an accelerometer (±16 *g*), gyroscope (±2,000 degrees/s), and magnetometer (±8 Gauss) and measures 43.7 × 39.7 × 13.7 mm (length × width × height) and weighs approximately 25 g. Participants completed two trials of a self-paced 20-m walk test at baseline. Gait parameters available from the MOST database included spatiotemporal and other features that were extracted from the raw sensor data at the MOST Accelerometry Center using published algorithms.[13–20] For each gait parameter, we used the mean of the two trials in the analyses.

Clinical and demographic features. We selected baseline measures that have been reported to be related to worsening pain or function[21–24] because many of these have also been associated with cartilage damage. These measures included age, sex, height, weight, depressive symptoms, race, clinic site, and measures of radiographic OA for each knee. At one of the two MOST clinic sites, Iowa, almost all participants are White, so covariate analyses of race are inherently confounded by clinic site. Rather than including these variables separately in the analysis, we created a variable that we shall call clinic/race with levels: Alabama (White), Alabama (Black), Alabama (other), and Iowa. Depressive symptoms were assessed using the Center for Epidemiological Studies Depression Scale (CES-D).[25] Radiographic OA was defined as KLG ≥2 on a weight-bearing posterior–anterior fixed flexion or weight-bearing lateral radiograph. Knee magnetic resonance imaging (MRI) was not acquired in knees that in previous examinations had KLG 3 or 4 or those with total knee

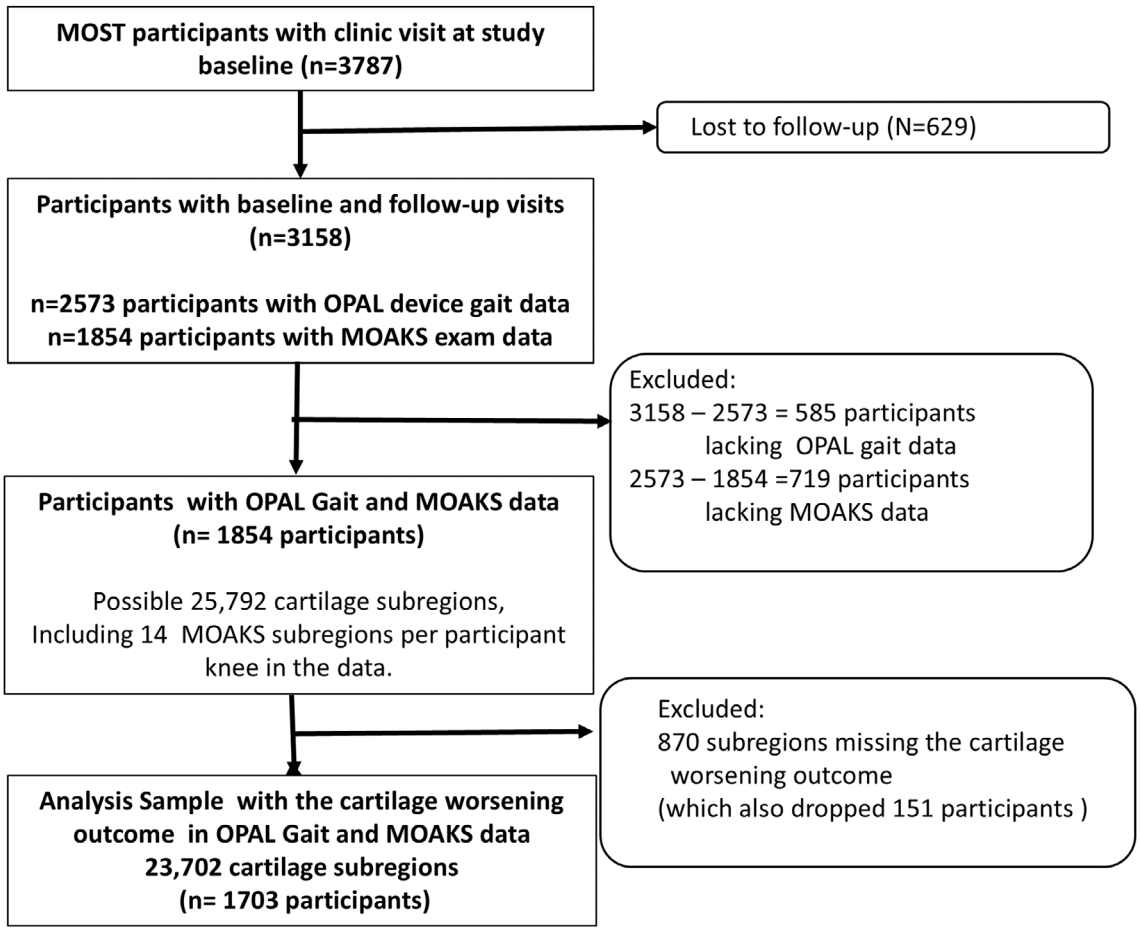

Figure 1. Cohort selection. Selection of the analysis cohort is shown. exam, examination; MOAKS, MRI Osteoarthritis Knee Score; MOST, Multicenter Osteoarthritis Study.

replacements, and therefore, we could not assess cartilage damage worsening in these knees, and they were excluded from the current analysis.

Outcome. Two experienced musculoskeletal radiologists (AG and FWR) completed paired readings of knee MRI acquired at the baseline and two-year follow-up visits. We read one randomly selected knee from each participant, and the same knee was read at two-year follow-up for each participant. The MOAKS[26,27] was used to assess the severity of cartilage damage in each of the 14 subregions of the knee,[12] assessing both area and depth of cartilage damage. For each subregion, a baseline cartilage score was created from the sum of the area and depth baseline scores. To increase sensitivity to change, within-grade changes were applied that fulfill definitive visual change but not the criteria of a full-grade increase. Within-grade changes have been shown to be clinically valid and are associated with quantitative cartilage loss.[26,27] We defined cartilage damage worsening as any within-grade or full-grade increase in MOAKS cartilage score in a subregion in a knee over two years. This change could occur in subregions with or without cartilage damage at baseline. The analysis data set includes one record per subregion per knee; as defined above, only one knee per participant is included.

Machine-learning process for variable selection. Our goal was to identify gait characteristics associated with cartilage damage worsening within a subregion over two years. We examined each subregion as our analytic outcome. We examined correlations among all gait measures and for all near perfect correlations (>0.85), we retained one feature. Gait features retained in the model are shown in Table 1. Other variables included in the model were age, height, weight, CES-D, sex, race/clinic site, and radiographic OA status. After excluding subregions for which we could not create the outcome cartilage damage worsening, we imputed missing values for other data variables using multiple imputation by chained equations.[28]

We used an ensemble machine-learning technique ("superlearning") as a feature reduction approach to identify important variables associated with the presence/absence of the outcome. Superlearner uses a multifold cross-validation to select the optimal combination of algorithms that theoretically achieves an

Table 1. Gait variables from the MOST database included in the machine-learning models*

Variable	Explanation
Gait speed (m/s)	Total distance walked/total walking time.
Cadence (steps/min)	Number of steps taken per minute.[13]
Step length (m)	Total distance walked/number of steps taken.
Stride time (s)	Mean duration of gait cycle (ie, stride).
Leg time CV (per knee)	CV of the leg mean time.
Leg stance percent (per knee)	The stance is the period of the gait cycle during which the foot remains in contact with the ground, typically ~60% of the gait cycle.
Leg stance time (s; per knee)	The stance is the period of the gait cycle during which the foot remains in contact with the ground.
Leg swing time (s; per knee)	The swing phase is the period of the gait cycle during which the foot is not in contact with the ground.
Step symmetry	Expresses the symmetry of the acceleration between left and right limbs.[14]
Gait asymmetry	Calculated as 100* | ln(right swing time/left swing time)|. Values of 0.0 reflect perfect symmetry, and higher values reflect greater degrees of asymmetry, with a range of 0 to 57.[15]
Stride time CV	Reflects the magnitude of the stride-to-stride variability of stride time.[16]
Step regularity	Measure of regularity of the vertical acceleration measured from the trunk sensor between consecutive steps using autocorrelation. Low step regularity indicates that there is a low regularity among steps or a systematic asymmetry between left and right leg.[14]
Stride regularity	Expresses the regularity of the vertical acceleration measured from the trunk sensor between consecutive strides using autocorrelation,[14] in which a stride refers to the whole gait cycle (left and right steps).
Sample entropy	Entropy is a measure that quantifies regularity in time series: the more predictable and less complex a series is, the lower the entropy value.[17]
PCI	A measure of bilateral coordination of gait assessed by quantifying the phase relationship between the step timing of the left and right legs.[18]
Phase CV	CV of the phase.[18]
Phase absolute difference	Measure of how close the phases are to 180°.
Dominant frequency	Frequency with the largest amplitude in the power spectral density in the 0.5–3 Hz band of the vertical acceleration signal from the trunk sensor.[19]
Amplitude of dominant frequency	The peak of the power spectral density of the vertical acceleration signal from the trunk sensor.[19]
Width of dominant frequency	Width of the peak at half of its peak amplitude.[19]
Lyapunov exponent (short divergent)	Estimate of local stability during gait.[20]

* CV, coefficient of variation; MOST, Multicenter Osteoarthritis Study; PCI, Phase Coordination Index.

accuracy superior to any single machine-learning method.[29,30] Our superlearner included a stacked ensemble of algorithms appropriate for binomial outcomes: logistic regression (fast generalized linear model library using Cholesky method); xgboost (extreme gradient boosting that can capture complex nonlinear relationships, with learning rate controlled to 0.3 to prevent overfitting, limit to 20 fitting iterations, max depth of tree = 6); and random forest (ranger: Fast(er) Random Forests, with one thread and max 500 trees to forest, stored for prediction). We randomly split the data into 70% training (development) and 30% test (evaluation) sets, ensuring the same proportion of the outcome in both data sets and ensuring that data from any participant's knee were only included in either training or test data set.[28] We trained candidate learners using the training data set and used five-fold cross-validation. As part of the superlearner approach, these training learners (the same set of learners; eg, xgboost) were used for predictions on out-of-fold samples and to develop a meta-learner that optimized the contribution of each individual learner (eg, determined the weight of xgboost contributions to the overall superlearner). The meta-learner was then applied to the held-out test set, and its performance was assessed using the area under the receiver operating characteristic curve (AUC). To increase robustness, the random data split and model training and testing were repeated 100 times. We report the median and the 2.5th and 97.5th percentile of the AUC across the 100 iterations.

During each iteration, a metric is calculated for each feature that compares the risk without the feature to the risk when it is included ("least squares error" on each feature). A higher difference value corresponds to a more important feature. We defined influential variables as the 10 variables most frequently identified as contributing to the outcome prediction across the 100 iterations.[11,22,31]

As an additional test of feature selection, we ran two additional stacked regression analyses that included least absolute shrinkage and selection operator (LASSO) in the stack: one adding LASSO to the current stack (ie, logistic regression, xgboost, random forest) and one with only LASSO and xgboost in the stack.

Statistical analyses. We used G-computation to evaluate causal risk differences for these "influential" variables with the outcome worsening cartilage damage by subregion over two years.[32] Confidence intervals were calculated with 1,000 bootstrap samples. We included the 10 influential variables selected from machine learning as exposures while adjusting for potential confounders such as age, race, and sex irrespective of whether they were identified as an influential predictor in the machine-learning process.

In G-computation analyses, all predictors were modeled as categorical measures. In the absence of any clinically meaningful thresholds, the predictors on the continuous scale (gait measures, weight) were modeled as tertiles with the lowest tertile as

the reference. For structural measures (baseline cartilage damage, KLG), no damage was the reference. For pain, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain <3 was the reference. For race/site, the Iowa site category was the reference. We also adjusted for the baseline value of the subregion's MOAKS cartilage score.

We evaluated the model variables for collinearity by reviewing correlations, tolerance values, and variance inflation factors. We then adjusted the set of model variables as necessary to deal with collinearity by choosing variables to drop from the model and then retesting the collinearity diagnostics. As a sensitivity analysis, we ran the same G-computation model in a subset including only subregions with cartilage damage at baseline. Based on results of the G-computation analyses, we tested for nonlinear associations using spline analyses for continuous gait variables.

Analyses were performed using SAS software version 9.4 and R version 4.0.2, with alpha level for significance set at <0.05 two sided. We followed the published guidelines for reporting on observational cohort studies from Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). The STROBE checklist is included with the supplementary materials.

Data from 25,792 subregions across knees of 1,854 participants were initially included, which were reduced to 23,702 subregions across 1,703 participants after subregions lacking the outcome were excluded (Figure 1; Table 2). No observations were missing gait data. Multiple imputation imputed data for 417 variable slots across 23,702 subregion observations and 34 variables (0.05% of data imputed). Variables with imputed data included age, sex, race, WOMAC pain (both knees), KLG (both knees), and baseline cartilage score. The most missing data were present in the KLGs for the contralateral knee.

Data from the MOST are publicly available at the National Institute on Aging, Aging Research Biobank, at https://agingresearchbiobank.nia.nih.gov/studies/most/. The MOST has been conducted in accordance with US Department of Health and Human Services Protection of Human Subjects regulations (45 CFR part 46) and the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA) of 1996. Research data and image sets are de-identified in accordance with regulation 45 CFR 164.514(e). Study participants were recruited under local institutional review board (IRB) approval through the collaborating institutions: Boston University IRB FWA00000301, The University of Iowa IRB FWA00003007, The University of Alabama at Birmingham IRB FWA00005960, and University of California, San Francisco, IRB FWA00000068.

All participants underwent an informed consent process approved by the IRB committee on human research at the participating institutions as listed above. Written consent is obtained from all participants at the beginning of a clinic visit at the clinical center. The consent covers all data collection scheduled for each grant cycle. Verbal consent is obtained for telephone interviews. Participants give written permission for clinical centers to obtain medical records needed for documentation of joint replacement surgery. There were no participants younger than 16 years of age.

Table 2. Participant characteristics (n = 1,703)*

Characteristic	Value, mean (±SD)
Age, y	61.2 (±9.3)
Female, %	56
Race, %	
White or Caucasian	83
Black or African American	15
Asian	1
American Indian or Alaskan Native	0.1
More than one race	0.8
Other	1
Clinic site, %	
University of Alabama	43
Iowa	57
Height, m	1.7 (±0.09)
Weight, pounds	181.8 (±37.4)
Depressive symptoms (CES-D), % > 16	10.7
Kellgren–Lawrence grade, % knees	
0	66
1	19
2	9
3	4
4	2
Cartilage damage score, baseline per subregion in knee (continuous)	0.48 (±1.04)
Lyapunov exponent (short divergence)	1.23 (±0.15)
Average step length, m	0.66 (±0.08)
Width of the dominant frequency in the power spectrum	0.69 (±0.07)
Phase Coordination Index	4.43 (±2.78)
WOMAC pain (ipsilateral knee)	1.82 (±2.767)
Step regularity	0.83 (±0.09)

* Table values are per-person estimates unless otherwise noted in the description. CES-D, Center for Epidemiological Studies Depression Scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

RESULTS

After removing subregions without the outcome, data from 23,702 subregions across knees of 1,703 participants were included (Figure 1; Table 2). In general, participants were elderly (mean ± SD age 61.2 ± 9.2 years), female (56% women), predominantly White (83%), and overweight (mean body mass index 28.7 ± 5.0). At baseline, approximately 13.3% of the participants had radiographic knee OA in at least one knee and 79% had cartilage damage in at least one subregion in a knee (MOAKS of subregion baseline ≥2 on a scale of 0–6). At two years, 46% of participants had worsening cartilage damage in at least one knee subregion, whereas 5.6% of all subregions across all knees exhibited worsening damage. Worsening cartilage damage differed across knee compartments: 7.7% of subregions in patellofemoral compartments showed worsening, compared to 5.6%

of subregions in medial tibiofemoral compartments, and 3.5% of subregions in lateral tibiofemoral compartments. Across 100 iterations, the median AUC (2.5th and 97.5th percentile) on the test sets was 0.81 (0.79, 0.83), sensitivity was 0.07 (0.04, 0.12), and specificity was 0.99 (0.98, 0.99). The 10 features that most frequently appeared as top contributors to prediction across 100 iterations (and frequency of appearance) were baseline cartilage MOAKS (100), KLG (44), Lyapunov exponent (39), weight (37), step length (36), Phase Coordination Index (35), width of dominant frequency (35), clinic/race (35), step regularity (33), and WOMAC pain in the ipsilateral knee (33). As noted in the Participants and Methods section, in addition to these variables, we added adjustment for age and sex in G-computation models.

Risk differences from the G-computation analyses can be interpreted as the difference in risk of outcome of worsening cartilage damage per tertile or unit increase in the risk factor variable compared to the referent category per 1,000 subregions.

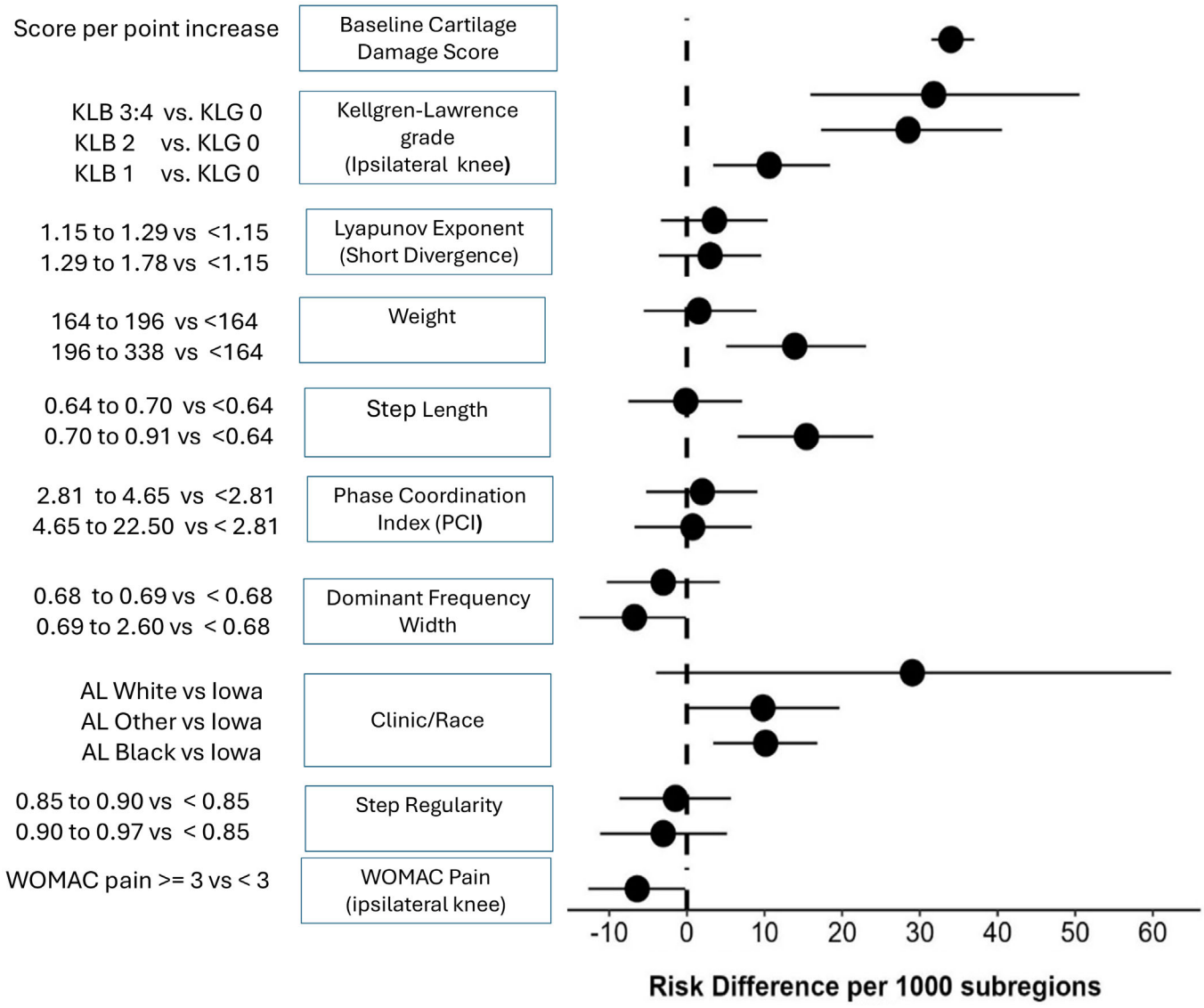

Figure 2. Risk differences (G-computation) for cartilage damage with variables chosen as influential in the machine-learning process. The figure displays the number of subregions per 1,000 that would be affected by this risk factor. The baseline cartilage score was modeled as a continuous variable. Kellgren–Lawrence grade in the ipsilateral knee was categorical (reference group 0 had no osteoarthritis). Lyapunov exponent (short divergence) was modeled in tertiles (reference was the lowest tertile). Weight was modeled in tertiles (reference was the lowest tertile). Step length was modeled in tertiles (reference was the lowest tertile). Phase Coordination Index was modeled in tertiles (reference was the lowest tertile). Dominant frequency width was modeled in tertiles (reference was the lowest tertile). Clinic/race was the combination clinic and race variable (reference was site Iowa White). Step regularity was modeled in tertiles (reference was the lowest tertile). Western Ontario and McMaster Universities Osteoarthritis Index pain (ipsilateral knee) was modeled as the dichotomous variable (reference was score <3). AL, Alabama; KLG, Kellgren–Lawrence grade; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80018/abstract.

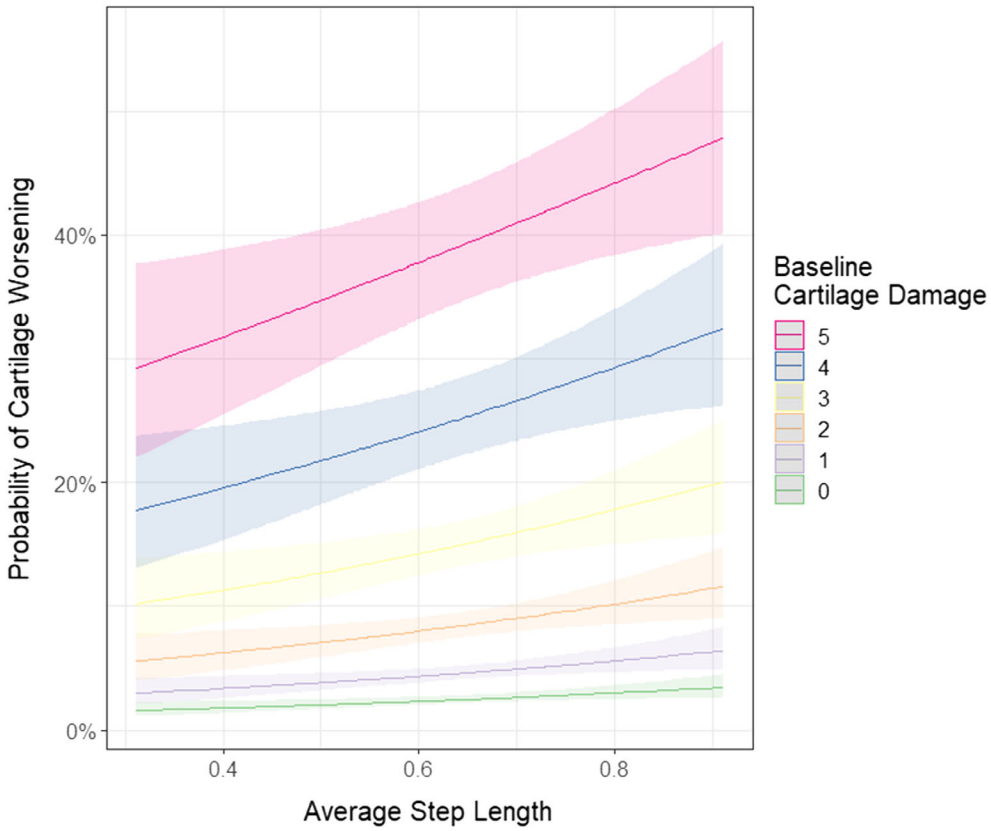

Figure 3. Predicted probability of worsening cartilage damage by average step length, grouped by baseline cartilage damage.

G-computation showed that longer step length was associated with an increased risk of worsening cartilage damage over two years (Figure 2). Specifically, per 1,000 subregions, 15 more subregions are expected to show worsened damage over two years in the highest tertile of step length compared to the lowest (number of knees needed to treat = 4.8, assuming one subregion worsening per knee). However, we did not see an increased risk for individuals in the middle tertile of step length compared to the lowest tertile, nor did we detect nonlinearity from spline analysis with average step length and worsening cartilage damage. Also associated with worsening cartilage damage were baseline cartilage damage in that subregion, radiographic OA status in the knee, weight, clinic/race, and WOMAC pain in the knee. Sensitivity analysis in a subset of participants with baseline cartilage damage (n = 4,948 subregions) showed larger risk differences for step length compared to subregions with no baseline damage (Figure 3).

In the additional tests of feature selection with alternative algorithms, the results for top 10 features were similar to those reported above. In particular, for average step length, for the first set of alternative algorithms (LASSO, logistic, xgboost, and random forests), average step length appeared in the top 10 features 37 times; for the second set of alternative algorithms (LASSO and xgboost only), average step length appeared 35 times.

DISCUSSION

Although demographic and/or lifestyle factors continue to be the strongest overall predictors for cartilage worsening, we did find associations with some gait characteristics, the focus of this analysis. Our results suggest that longer step length, a modifiable characteristic of gait, is associated with worsening cartilage damage over two years in individuals with, or at risk for, knee OA, particularly in those with more baseline cartilage damage. Interventions to reduce step length could be considered to reduce risk of cartilage damage worsening. However, whether step length can be modified to an extent sufficient to reduce the risk of worsening cartilage damage needs further study.

Importantly, we previously identified longer step length as being a risk factor for worsening knee pain.[11] We have also previously reported that cartilage damage worsening is associated with worsening knee pain.[33] To our knowledge, this is the first time such a dual association of a modifiable gait parameter with multiple outcomes has been reported in a longitudinal study in OA. Focusing rehabilitation efforts on this or related parameters of gait may be effective not only in reducing pain and improving function, the most important symptoms to patients, but also for having favorable impacts on knee structure.

Longer step length has been associated with greater loading at the knee, which may explain the association[34,35] we found. Although measures of knee loading during walking (eg, peak knee adduction moment, knee adduction moment impulse, joint contact forces) are known to be related to greater risk of structural worsening of knee OA,[2] these measures may be challenging to modify with treatment, and their association with knee pain is inconsistent in the literature.[36] Our study provides important longitudinal evidence in support of a relation between easily measured gait features and changes in structural features of OA. Other evidence includes a study in which reducing stride length by 15% significantly reduced measures of knee loading in women with obesity and with healthy weight and increased the metabolic cost of walking.[37] As such, interventions to reduce step length may not only provide biomechanical benefits in terms of reducing joint loading but could elicit a metabolic response consistent with engaging in exercise, both important considerations for people with knee OA. Interventions to modulate step length typically involve cuing strategies that show the targeted step length by visualizing footprints on the ground.[38–40] In a study of healthy individuals, shortening step/stride length was found to be most effective at reducing measures of knee loading.[41] However, in people with knee OA, reducing stride length was effective at reducing measures of knee loading but not in everyone.[42] In another study of knee OA, they found no longitudinal association of average step length with structural damage in the patellofemoral joint, as measured by four Whole-Organ Magnetic Resonance Imaging Score subregions.[43] In our study, the association of longer step length with worsening cartilage damage was only seen for those in the highest tertile compared to the lowest tertile; in knees with low or no baseline cartilage damage, longer step length may not predict cartilage damage worsening (as shown in Figure 3). Hence, it is possible the interventions may be appropriate only for certain individuals and large reductions in step length may be needed for efficacy.

Other gait features, including nonlinear features of gait regularity and/or complexity, were identified as important contributors to predictors of worsening cartilage damage in the machine-learning model. However, these gait features were not found to be associated with the outcome in the explanatory G-computation models. Unlike the stacked regression model used in the machine-learning process to assess possible important predictors in a model in which all potential predictors are considered equally, with G-computation, we focus on each chosen predictor from the machine-learning process in a separate model. The individual G-computation model for each feature treats it as the "exposure" in a model, with G-computation creating its counterfactual setting for all data and adjusting for other important predictors in the model. This estimates an effect for worsening cartilage damage for that variable compared to its referent category. Given the expected interrelationships among gait features, our robust interpretable machine-learning approach was effective at identifying the gait feature most likely to be related to increased risk of worsening cartilage damage after accounting for other important gait features.

The effect size we report is small—only 15 of 1,000 subregions were affected by a potential reduction in step length. However, this association is likely underestimated given the mild disease in our sample. We note (see Figure 3) that the association is greater in knees with more cartilage damage at baseline. This is borne out in the sensitivity analysis among those with cartilage damage at baseline. Supplementary Figure 1 shows the probability of cartilage worsening by height-normalized step length.

Using superlearning, we observed an AUC of 0.81 for prediction of worsening cartilage damage over two years. These results are comparable to prior studies using machine learning or deep learning that report AUCs between 0.65 and 0.86 for prediction of structural worsening of knee OA over 12 to 48 months.[44–48] However, most prior studies relied on a radiographic definition of structural progression, whereas we used cartilage damage from MRI. Further, none of the prior studies included gait measures as predictors and mostly relied on image-based predictors, which would be expected to do well for predicting an image-based outcome. Our results are particularly compelling given the fact that a majority of our cohort had KLG <2. Our results are also substantially better than the AUC of 0.61 and 0.60 that we reported for prediction of worsening pain or worsening physical function using similar predictors and machine-learning approach.[11] To our knowledge, there are no currently available data sets that include the measures that we had available in MOST for an external validation of our machine-learning model. However, if this model could be validated, it could serve as a potential tool for identifying individuals who may be at risk of worsening cartilage damage and worse pain and/or physical function.

Limitations to our study should be considered. We cannot rule out residual confounding from other causes of cartilage damage. Other approaches to select influential predictors from machine learning and then determine associations may yield different results, although our process using superlearning and G-computation included multiple steps intended to maximize the robustness of our findings. Currently, there is no consensus in the field on these approaches, and our analysis could be considered exploratory. Wearable sensor–derived gait measures, although easy to measure in clinical settings, do not provide an estimate of knee joint loading. The two-year follow-up may limit the number of outcomes we detect for cartilage worsening. We did not have sufficient numbers of outcomes for a formal interaction test of step length by baseline cartilage damage, nor did we have enough compartment specific outcome events to produce credible subregional results. Finally, the thresholds for upper and lower tertiles of step length are specific to our cohort and may not translate to other cohorts or populations.

To conclude, we observed that greater step length was related to a greater risk of worsening cartilage damage over two

years, particularly in those with greater damage at baseline. Shortening step length may be associated with reduced risk of cartilage damage and worsening pain. This provides a possible avenue for gait related modifications that might be effective in knee OA.

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 Bacon 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. Vos T, Lim SS, Abbafati C, et al; GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020;396(10258):1204–1222. doi:https://doi.org/10.1016/S0140-6736(20)30925-9 Erratum in: *Lancet* 2020;396:1562.
2. D'Souza N, Charlton J, Grayson J, et al. Are biomechanics during gait associated with the structural disease onset and progression of lower limb osteoarthritis? A systematic review and meta-analysis. Osteoarthritis Cartilage 2022;30(3):381–394. doi:https://doi.org/10.1016/j.joca.2021.10.010
3. 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. doi:https://doi.org/10.1002/art.41142 Erratum in: *Arthritis Rheumatol* 2021;73:799.
4. Kobsar D, Masood Z, Khan H, et al. Wearable inertial sensors for gait analysis in adults with osteoarthritis—a scoping review. Sensors (Basel) 2020;20(24):7143. doi:https://doi.org/10.3390/s20247143
5. Rose MJ, Costello KE, Eigenbrot S, et al. Inertial measurement units and application for remote health care in hip and knee osteoarthritis: narrative review. JMIR Rehabil Assist Technol 2022;9(2):e33521. doi:https://doi.org/10.2196/33521
6. Boekesteijn RJ, van Gerven J, Geurts ACH, et al. Objective gait assessment in individuals with knee osteoarthritis using inertial sensors: a systematic review and meta-analysis. Gait Posture 2022;98:109–120. doi:https://doi.org/10.1016/j.gaitpost.2022.09.002
7. Gao C, Sun H, Wang T, et al. Model-based and model-free machine learning techniques for diagnostic prediction and classification of clinical outcomes in Parkinson's disease. Sci Rep 2018;8(1):7129. doi:https://doi.org/10.1038/s41598-018-24783-4
8. Ricotti V, Kadirvelu B, Selby V, et al. Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. Nat Med 2023;29(1):95–103. doi:https://doi.org/10.1038/s41591-022-02045-1
9. Bernad-Elazari H, Herman T, Mirelman A, et al. Objective characterization of daily living transitions in patients with Parkinson's disease using a single body-fixed sensor. J Neurol 2016;263(8):1544–1551. doi:https://doi.org/10.1007/s00415-016-8164-6
10. Hannink J, Kautz T, Pasluosta CF, et al. Sensor-based gait parameter extraction with deep convolutional neural networks. IEEE J Biomed Health Inform 2017;21(1):85–93. doi:https://doi.org/10.1109/JBHI.2016.2636456
11. Bacon KL, Felson DT, Jafarzadeh SR, et al. Gait alterations and association with worsening knee pain and physical function: a machine learning approach with wearable sensors in the Multicenter Osteoarthritis Study. Arthritis Care Res (Hoboken) 2024;76(7):984–992. doi:https://doi.org/10.1002/acr.25327
12. Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011;19(8):990–1002. doi:https://doi.org/10.1016/j.joca.2011.05.004 Erratum in: *Osteoarthritis Cartilage* 2011;19:1168.
13. Slaght J, Sénéchal M, Hrubeniuk T, et al. Walking cadence to exercise at moderate intensity for adults: a systematic review. J Sports Med (Hindawi Publ Corp) 2017;2017:4641203.
14. Moe-Nilssen R, Helbostad JL. Estimation of gait cycle characteristics by trunk accelerometry. J Biomech 2004;37(1):121–126. doi:https://doi.org/10.1016/S0021-9290(03)00233-1
15. Yogev G, Plotnik M, Peretz C, et al. Gait asymmetry in patients with Parkinson's disease and elderly fallers: when does the bilateral coordination of gait require attention? Exp Brain Res 2007;177(3):336–346. doi:https://doi.org/10.1007/s00221-006-0676-3
16. Hausdorff JM. Gait dynamics in Parkinson's disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. Chaos 2009;19(2):026113. doi:https://doi.org/10.1063/1.3147408
17. Ramdani S, Seigle B, Lagarde J, et al. On the use of sample entropy to analyze human postural sway data. Med Eng Phys 2009;31(8):1023–1031. doi:https://doi.org/10.1016/j.medengphy.2009.06.004
18. Plotnik M, Giladi N, Hausdorff JM. A new measure for quantifying the bilateral coordination of human gait: effects of aging and Parkinson's disease. Exp Brain Res 2007;181(4):561–570. doi:https://doi.org/10.1007/s00221-007-0955-7
19. Weiss A, Sharifi S, Plotnik M, et al. Toward automated, at-home assessment of mobility among patients with Parkinson disease, using a body-worn accelerometer. Neurorehabil Neural Repair 2011;25(9):810–818. doi:https://doi.org/10.1177/1545968311424869
20. Bruijn SM, Bregman DJ, Meijer OG, et al. Maximum Lyapunov exponents as predictors of global gait stability: a modelling approach. Med Eng Phys 2012;34(4):428–436. doi:https://doi.org/10.1016/j.medengphy.2011.07.024
21. Felson DT, Niu J, Quinn EK, et al. Multiple nonspecific sites of joint pain outside the knees develop in persons with knee pain. Arthritis Rheumatol 2017;69(2):335–342. doi:https://doi.org/10.1002/art.39848
22. Bacon KL, Felson DT, Jafarzadeh SR, et al; Multicenter Osteoarthritis Study Investigators. Relation of gait measures with mild unilateral knee pain during walking using machine learning. Sci Rep 2022;12(1):22200. doi:https://doi.org/10.1038/s41598-022-21142-2
23. Bacon KL, Segal NA, Øiestad BE, et al. Concurrent change in quadriceps strength and physical function over five years in The Multicenter Osteoarthritis Study. Arthritis Care Res (Hoboken) 2019;71:1044–1051.
24. de Rooij M, van der Leeden M, Heymans MW, et al. Prognosis of pain and physical functioning in patients with knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2016;68(4):481–492. doi:https://doi.org/10.1002/acr.22693
25. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385–401. doi:https://doi.org/10.1177/014662167700100306
26. Roemer F, Maschek S, Wisser A, et al. Worsening of articular tissue damage as defined by semi-quantitative MRI is associated with concurrent quantitative cartilage loss over 24 months. Cartilage 2023;14(1):39–47. doi:https://doi.org/10.1177/19476035221147677
27. Roemer FW, Nevitt MC, Felson DT, et al. Predictive validity of within-grade scoring of longitudinal changes of MRI-based cartilage

morphology and bone marrow lesion assessment in the tibio-femoral joint-the MOST study. Osteoarthritis Cartilage 2012;20(11):1391–1398. doi:https://doi.org/10.1016/j.joca.2012.07.012

28. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1–67. doi:https://doi.org/10.18637/jss.v045.i03
29. Coyle JR, Hejazi NS, Malenica I, et al. sl3: Pipelines for Machine Learning and Super Learning. Version 1.4.2. Zenodo. December 23, 2021. Accessed March 28, 2023. https://doi.org/10.5281/zenodo.1342293
30. Wolpert DH. Stacked generalization. Neural Netw 1992;5(2):241–259. doi:https://doi.org/10.1016/S0893-6080(05)80023-1
31. Costello KE, Felson DT, Jafarzadeh SR, et al. Gait, physical activity and tibiofemoral cartilage damage: a longitudinal machine learning analysis in the Multicenter Osteoarthritis Study. Br J Sports Med 2023;57(16):1018–1024. doi:https://doi.org/10.1136/bjsports-2022-106142
32. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol 2011;173(7):731–738. doi:https://doi.org/10.1093/aje/kwq472
33. Bacon K, LaValley MP, Jafarzadeh SR, et al. Does cartilage loss cause pain in osteoarthritis and if so, how much? Ann Rheum Dis 2020;79(8):1105–1110. doi:https://doi.org/10.1136/annrheumdis-2020-217363
34. DeVita P, Rider P, Hortobágyi T. Reductions in knee joint forces with weight loss are attenuated by gait adaptations in class III obesity. Gait Posture 2016;45:25–30. doi:https://doi.org/10.1016/j.gaitpost.2015.12.040
35. Lyle MA, Jensen JC, Hunnicutt JL, et al. Associations of strength and spatiotemporal gait variables with knee loading during gait after anterior cruciate ligament reconstruction. J Athl Train 2022;57(2):158–164. doi:https://doi.org/10.4085/1062-6050-0186.21
36. Hutchison L, Grayson J, Hiller C, et al. Relationship between knee biomechanics and pain in people with knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2023;75(6):1351–1361. doi:https://doi.org/10.1002/acr.25001
37. Russell EM, Braun B, Hamill J. Does stride length influence metabolic cost and biomechanical risk factors for knee osteoarthritis in obese women? Clin Biomech (Bristol) 2010;25(5):438–443. doi:https://doi.org/10.1016/j.clinbiomech.2010.01.016
38. Lim YP, Lin Y-C, Pandy MG. Effects of step length and step frequency on lower-limb muscle function in human gait. J Biomech 2017;57:1–7. doi:https://doi.org/10.1016/j.jbiomech.2017.03.004
39. Buddhadev HH, Smiley AL, Martin PE. Effects of age, speed, and step length on lower extremity net joint moments and powers during walking. Hum Mov Sci 2020;71:102611. doi:https://doi.org/10.1016/j.humov.2020.102611
40. Gill N, O'Leary T, Roberts A, et al. Enforcing walking speed and step-length affects joint kinematics and kinetics in male and female healthy adults. Gait Posture 2023;103:223–228. doi:https://doi.org/10.1016/j.gaitpost.2023.05.025
41. Edd SN, Bennour S, Ulrich B, et al. Modifying stride length in isolation and in combination with foot progression angle and step width can improve knee kinetics related to osteoarthritis; a preliminary study in healthy subjects. J Biomech Eng 2020;142(7):074505. doi:https://doi.org/10.1115/1.4046713
42. Ulrich B, Pereira LC, Jolles BM, et al. Walking with shorter stride length could improve knee kinetics of patients with medial knee osteoarthritis. J Biomech 2023;147:111449. doi:https://doi.org/10.1016/j.jbiomech.2023.111449
43. Stefanik JJ, Gross KD, Guermazi A, et al. Relation of step length to magnetic resonance imaging–detected structural damage in the patellofemoral joint: the Multicenter Osteoarthritis Study. Arthritis Care Res (Hoboken) 2016;68(6):776–783. doi:https://doi.org/10.1002/acr.22738
44. Yin R, Chen H, Tao T, et al. Expanding from unilateral to bilateral: a robust deep learning-based approach for predicting radiographic osteoarthritis progression. Osteoarthritis Cartilage 2024;32(3):338–347. doi:https://doi.org/10.1016/j.joca.2023.11.022
45. Schiratti JB, Dubois R, Herent P, et al. A deep learning method for predicting knee osteoarthritis radiographic progression from MRI. Arthritis Res Ther 2021;23(1):262. doi:https://doi.org/10.1186/s13075-021-02634-4
46. Guan B, Liu F, Haj-Mirzaian A, et al. Deep learning risk assessment models for predicting progression of radiographic medial joint space loss over a 48-month follow-up period. Osteoarthritis Cartilage 2020;28(4):428–437. doi:https://doi.org/10.1016/j.joca.2020.01.010
47. Du Y, Almajalid R, Shan J, et al. A novel method to predict knee osteoarthritis progression on MRI using machine learning methods. IEEE Trans Nanobioscience 2018;17(3):228–236. doi:https://doi.org/10.1109/TNB.2018.2840082
48. Halilaj E, Le Y, Hicks JL, et al. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2018;26(12):1643–1650. doi:https://doi.org/10.1016/j.joca.2018.08.003

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1239–1246
DOI 10.1002/acr.80021
© 2026 The Author(s). *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Moderators of the Effects of Exercise and Manual Therapy for People With Knee and Hip Osteoarthritis: A Secondary Analysis of a Randomized Clinical Trial

Daniel Cury Ribeiro, PT, MSc, PhD,[1] Sarah E. Lamb, PT, DPhil,[2] and J. Haxby Abbott, DPT, FNZCP, PhD[3]

Objective. This study aimed to investigate potential moderators influencing the effects of manual therapy and exercise therapy on pain and functional outcomes in individuals with knee and/or hip osteoarthritis, using data from the MOA trial. This is a secondary analysis of data from the MOA trial that compares the clinical effectiveness of manual therapy and/or exercise therapy in addition to usual care for patients with hip and/or knee osteoarthritis.

Methods. A total of 206 participants were analyzed. The primary outcome measure was the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) composite score after one year. The secondary outcome measures were WOMAC pain and function scores. We used linear regression models for assessing whether the effect of randomized interventions on pain and function was moderated by body mass index (BMI), pain self-efficacy, quadriceps strength, mental health, and education.

Results. BMI moderated the treatment effects of manual therapy interventions on WOMAC composite (β –4.1, 95% confidence interval [CI] –6.6 to –1.7), function (β –3.3, 95% CI –5.1 to –1.6), and pain scores (β –0.6, 95% CI –1.3 to 0.0) when compared to usual care. A negative β reflects an improvement in outcome associated with each one-unit increase in BMI. Mental health moderated (β 16.8, 95% CI 2.2–31.4) treatment effects of manual therapy interventions when compared to usual care. No other moderation effect was identified.

Conclusion. Our findings suggest manual therapy may prove to be a more suitable treatment for people with higher BMI and should undergo further trials. Future trials should continue to explore which variables moderate treatment effects when testing effectiveness of nonsurgical interventions in patients with hip or knee osteoarthritis.

INTRODUCTION

Nonsurgical interventions such as exercise therapy (including strengthening and aerobic exercises) are recommended for people with hip and knee osteoarthritis.[1] According to the Osteoarthritis Research Society International guidelines, management should include education about the condition and structured land-based exercises with or without dietary weight management.[1] Tailoring interventions to people with hip or knee osteoarthritis could improve health outcomes and optimize resource allocations within health care services.

Previous reviews have examined treatment-effect moderators for people with hip or knee osteoarthritis exposed to exercise therapy.[2,3] Findings from one review[3] suggest there is little evidence that presence, number and type of comorbidity, psychosocial factors (eg, anxiety or depression), obesity, or quadriceps strength moderate the effects of exercise therapy.[3] A recent systematic review and individual participant data meta-analysis found pain and function at baseline moderated treatment effects of exercise therapy.[2] That review did not explore treatment-effect moderators of manual therapy interventions. So far, little research has explored the moderators of exercise

ANZCTR no.: ACTRN12608000130369.
The Health Research Council – New Zealand and the New Zealand Lottery Grants Board had no role in the design of the project and the draft and interpretation of the findings.
The research was conducted during the tenure of The Sir Charles Hercus Health Research Fellowship of the Health Research Council of New Zealand (grant 18/111, awarded to Dr Ribeiro). The MOA Trial was funded by research contracts from the Health Research Council of New Zealand (HRC 07/199 and 07/200) and the New Zealand Lottery Grants Board (MR212664).
[1]Curtin University, Perth, Western Australia, Australia; [2]University of Exeter, St Luke's Campus, Exeter, United Kingdom; [3]University of Otago, Dunedin, New Zealand.

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.80021).
Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80021.
Address correspondence via email to Daniel Cury Ribeiro, PT, MSc, PhD, at daniel.curyribeiro@curtin.edu.au.
Submitted for publication May 2, 2025; accepted in revised form February 6, 2026.

SIGNIFICANCE & INNOVATIONS

- Body mass index (BMI) moderated the effects of manual therapy on Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) composite and function scores in patients with hip or knee osteoarthritis but did not moderate the effects of exercise therapy.
- Mental health moderated the effects of manual therapy on WOMAC function scores.
- Manual therapy may be a more suitable treatment for patients with hip or knee osteoarthritis with higher BMI.
- Confirmatory trials are needed to estimate treatment effects of manual therapy in patients with hip or knee osteoarthritis and higher BMI.

and manual therapy interventions for patients with hip or knee osteoarthritis.

The MOA trial[4] compared manual physiotherapy and/or exercise physiotherapy in addition to usual care for people with osteoarthritis of the hip or knee and found that exercise or manual therapy were more effective than usual care and that effects were maintained up to 12 months of follow-up. The MOA trial also reported no added benefit when combining exercise and manual therapy compared to usual care. This study aimed to investigate potential moderators influencing the effects of manual therapy and exercise therapy on pain and functional outcomes in individuals with knee and/or hip osteoarthritis using data from the MOA trial.

METHODS

Design. This is a secondary analysis of data from the MOA trial that compared the clinical effectiveness of manual therapy and/or exercise therapy in addition to usual care for patients with hip and/or knee osteoarthritis.[4] The MOA trial is a 2 × 2 factorial trial with a one-year follow-up. The trial was prospectively registered (ANZCTR no.: ACTRN12608000130369) and was approved by the Lower South Regional Ethics Committee (New Zealand Ministry of Health, reference: LRS/07/11/044). The protocol,[5] the outcome evaluation study,[4,6] and the mediation analyses[7] studies have already been published.

Study population. Two hundred six participants with hip and/or knee osteoarthritis took part in the original trial.[4] To be included, participants had to meet the American College of Rheumatology clinical criteria (ie, no radiographic criteria) for hip or knee osteoarthritis.[8–10] Of those participants, a total of 44 (21%) participants had arthroplasty of the index hip or knee during the follow-up period. Before taking part in the study, participants signed an informed consent. The rights of participants were protected.

Interventions. Participants were randomly allocated to one of the following interventions: no trial physiotherapy (n = 51), manual physiotherapy (n = 54), multimodal exercise physiotherapy (n = 51), or combined exercise and manual physiotherapy (n = 50).

1. Usual care only: participants allocated to usual care received routine care offered by their general practitioner or other health care provider, without direction, restriction, or influence by the researchers.
2. Usual care plus manual therapy: participants allocated to this group received a standard protocol of manual therapy interventions that were individually tailored to each participant's clinical presentation and were prescribed home-based exercises (three times per week) to improve range of motion.
3. Usual care plus exercise therapy: participants allocated to this group received a multimodal, supervised exercise program that included warm-up and aerobic exercises, strengthening and stretching exercises, and neuromuscular control exercises. Clinicians could prescribe additional exercises from a limited list of interventions based on findings from clinical assessment.
4. Usual care plus combined manual and exercise therapy: participants allocated to this group received a combination of interventions as described for the manual therapy and exercise groups.

Outcome measures. The primary outcome measure in the trial was the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) composite score after one year. Participants completed the questionnaire at baseline and at 9-week, 6-month, and 12-month follow-ups. We calculated the 12-month change. The MOA trial was powered to detect a minimum clinically important difference in the WOMAC score of 28 units (0–240 scale).[5] We also considered the following subscales as secondary outcome measures: WOMAC function scores and WOMAC pain scores.

Putative moderators. We selected moderators based on data available from the MOA trial by previous systematic reviews assessing moderators of treatment effect for patients with hip or knee osteoarthritis.[2,3] We considered as a moderator measures that described characteristics of participants and that were measured at baseline and, whenever possible, before randomization.[11] Based on that, the following putative moderators were included in the analyses: body mass index (BMI), pain self-efficacy, quadriceps strength, mental health, and education (Table 1).

BMI was calculated as weight in kg divided by height in m^2 for each participant. Pain self-efficacy was measured using scores from two questions from the Pain Belief Screening Instrument[15,16] (Table S1). Scores for each question range from 0 (not

Table 1. Putative moderators and their respective rationale

Moderator	Rationale
Body mass index	Based on the evidence that obesity moderated the effects of neuromuscular exercise on pain compared to strengthening exercise[12]
Pain self-efficacy	Based on evidence that self-efficacy moderated treatment effects on pain[13] or function scores[14]
Quadriceps strength	Based on conflicting findings in the literature[2,3]
Mental health	Based on previous studies exploring mental health as a moderator and the risk of false-negative findings by previous studies[3]
Education	Based on recommendations from a previous review suggesting trials to explore education level as a moderator[3]

at all confident) to 10 (very confident). Mental health was assessed using the two questions of a validated depression screening questionnaire[17] (more details described in the Supplementary Material). Participants were categorized into two groups: high or low risk of depression. The questions used to assess pain self-efficacy and mental health, possible answers, and scores are detailed in the Supplementary Material. Quadriceps strength was measured in kg using a handheld dynamometer. This questionnaire has a sensitivity of 97% (95% confidence interval [CI] 83%–99%) and a specificity of 67% (95% CI 62%–72%) for screening patients with depression.[17] Education was self-reported by questionnaire with the following options: less than three years of high school, three years or more of high school, some tertiary training, graduate degree, and postgraduate degree.

Time points. Participants' characteristics, outcome measures, and putative moderators were measured at baseline. WOMAC composite, function, and pain scores were also measured at 12 months' follow-up.

Sample size estimation. The MOA trial was designed to identify a minimum clinically important difference of 28 WOMAC points for each of the main effects (ie, manual therapy vs no manual therapy; exercise therapy vs no exercise therapy), assuming an SD of 50 points, Type I error rate of 0.05, and 95% power for the main effect analyses and 75% power for cell-by-cell analysis of active interventions versus control. For that, the sample size required was 180 participants. The authors considered a 10% dropout, and the minimum sample size was 200 participants. Given this is a post hoc secondary analysis of the MOA trial,[4] we are likely to be underpowered to detect treatment-effect moderation effects.[18,19]

Missing data. We assessed missing data patterns to determine the appropriate strategies for imputation of each variable. The analysis of patterns of missing data indicated missing data were missing completely at random.[20] We generated 36 imputations of missing values for each of variable with missing data. We checked estimates to confirm imputed values did not deviate significantly from observed data by using mean value, SD, and minimum and maximum values. We used the mi function in the statistical software package Stata (StataCorp)[21] for replacing missing data through multiple imputation.

Statistical analysis. We used descriptive statistics for summarizing participants' characteristics: with mean and SDs used for describing continuous variables or number and percentage used for describing categorical variables. We explored the rates of hip or knee arthroplasty using descriptive statistics. We used R Software[22] for conducting all analyses and set α at 0.05.

Arthroplasty was an unplanned, nonrandomized cointervention.[4] Surgery was a postrandomization event with a likely strong effect on outcome, and was likely associated with prognostic factors measured at baseline as well as other postrandomization factors (which may be unmeasured).[23–25] Given the effect of hip or knee arthroplasty on pain and function scores at 12 months' follow-up, we excluded participants who underwent surgery because this would distort assessment of the effects of their allocated intervention. We considered surgeries to be randomly distributed across the groups (ie, no suggestion of systematic difference) (Tables S2 and S3).

By restricting the analyses on a postrandomization event (ie, surgery) there is a risk of selection bias.[24,26] To address that risk of selection bias, we used inverse-probability weighting to adjust for prognostic factors that are associated with surgery.[23–25,27] Through inverse-probability weighting, we created a pseudopopulation in which everyone adheres to the allocated treatment, and through that, we adjusted treatment estimates to selection bias.[23–25,28]

Adherence adjustment using inverse-probability weighting. We used stabilized inverse-probability weights to adjust for selection bias. For that, we adopted the following steps[23–25]: (1) We estimated the inverse probability of adherence weights (ie, the probability of not receiving surgery) using the complete data set (n = 206) for each arm of the trial separately.[25] We used a logistic regression model to estimate the probability of adherence weights. We included the following parameters (measured at baseline) in the logistic regression model: age, WOMAC composite score, quadriceps muscle strength, BMI, and duration of symptoms. (2) We then conducted the moderation analyses using a weighted linear regression model with weights estimated during step 1. We present details for moderation analyses in the following section.

Moderation analysis. We used linear regression models for assessing whether the effect of randomized interventions on

pain and function was moderated by selected characteristics of participants at baseline. Our analyses estimated the incremental effects of manual therapy, exercise therapy, or combined manual and exercise therapy, relative to usual care alone. We performed linear regressions using the composite WOMAC score at the one-year follow-up as a dependent variable. We considered treatment allocation as an independent variable and age, BMI, number of years since symptom onset, quadriceps muscle strength, mental health (high risk of depression was used as the reference category), and pain self-efficacy as covariates. We included composite WOMAC scores at baseline as covariates in regression models. For assessing whether treatment effects were moderated by one of the putative moderators, we included a standard interaction term between the selected moderator and the exposure (ie, group allocation). We used standard diagnostic plots to assess regression assumptions of linearity and homoscedasticity.

We also visually assessed the interaction between exposure (ie, group allocation) and the potential moderator through plots of the WOMAC composite score at one year versus the potential moderator.

RESULTS

Sample characteristics. A total of 206 participants were enrolled in the trial, with 193 participants (93.2%) completing the one-year follow-up. The demographics of participants who did not receive surgery are presented in Table 2, and the demographics of participants who did receive surgery are presented in Table S2. Participants' mean age was 66.5 years (SD 9.5) and mean WOMAC composite score was 101 (SD 54.2) at baseline. Participants allocated to the usual care plus manual therapy group presented a slightly higher WOMAC composite score at

Table 2. Demographic characteristics of participants who did not receive total hip or knee replacement*

Variable	Usual care (n = 43)	Usual care plus manual therapy (n = 42)	Usual care plus exercise therapy (n = 40)	Usual care plus combined manual and exercise therapy (n = 37)
Demographics				
Age, mean (SD)	66.1 (10.4)	66.7 (11.1)	66.3 (7.6)	65.1 (9.3)
Female, n (%)	21 (48.8)	20 (47.6)	26 (65.0)	25 (67.6)
BMI, mean (SD)	29.8 (5.6)	29.1 (6.0)	29.3 (6.2)	30.7 (6.0)
BMI range	21.1–48.7	21.5–50.0	21.3–49.0	20.4–48.3
Clinical measurements				
WOMAC score (0–240),[a] mean (SD)	86.1 (51.6)	104.0 (56.4)	85.2 (56.0)	95.5 (50.1)
WOMAC pain (0–50),[a] mean (SD)	17.1 (10.8)	22.0 (11.9)	17.2 (11.0)	19.8 (9.7)
WOMAC function (0–170),[a] mean (SD)	60.0 (38.5)	72.2 (41.9)	59.4 (41.3)	65.7 (37.3)
Duration of symptoms, n (%)				
Less than 1 y	5 (11.6)	5 (11.9)	4 (10.0)	1 (2.7)
1–2 y	10 (23.3)	13 (31.0)	6 (15.0)	6 (16.2)
3–5 y	10 (23.3)	9 (21.4)	14 (35.0)	17 (45.9)
5–10 y	12 (27.9)	9 (21.4)	9 (22.5)	8 (21.6)
More than 10 y	6 (14.0)	6 (14.3)	7 (17.5)	5 (13.5)
Quadriceps muscle strength, kg/kg body mass, mean (SD)	0.21 (0.11)	0.21 (0.10)	0.21 (0.08)	0.20 (0.08)
Pain self-efficacy (0–20), mean (SD)	2.5 (2.9)	2.2 (2.6)	1.8 (2.4)	2.4 (2.5)
Mental health (low risk of depression), n (% of group)	18 (41.9)	19 (45.2)	18 (45.0)	15 (40.5)
Education, n (% of group)				
Less than 3 y of high school	16 (37.2)	16 (38.1)	10 (25.0)	13 (35.1)
3 y or more of high school	12 (27.9)	13 (31.0)	13 (32.5)	9 (24.3)
Some tertiary training	4 (9.3)	3 (7.1)	8 (20.0)	5 (13.5)
Graduate degree	5 (11.6)	8 (19.0)	7 (17.5)	7 (18.9)

* BMI, body mass index; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.
[a] Lower scores represent less pain, stiffness, and disability.

baseline (Table 2). Overall, the sample consisted of middle-aged and overweight adults.

The numbers of participants receiving arthroplasty in each group were as follows: no trial physiotherapy, 8 of 51 (15%); manual physiotherapy, 12 of 54 (22%); multimodal exercise physiotherapy, 11 of 51 (21%); and combined exercise and manual physiotherapy, 13 of 50 (26%). The numbers were not significantly different (P = 0.650 for chi-square difference in proportions). Participants who underwent surgery presented higher WOMAC composite, pain, and function scores at baseline (ie, worse clinical outcomes) when compared to those who did not undergo surgery (Table S3). The characteristics of participants receiving arthroplasty during the trial were mostly similar across the groups (Table S4). The characteristics of participants receiving arthroplasty, by joint site affected, are presented in Table S5, and the demographic characteristics at baseline for participants receiving arthroplasty, by sex, are presented in Table S6. There was a small imbalance between groups for WOMAC composite scores: participants allocated to the usual care plus manual therapy group presented higher WOMAC composite, pain, and function scores (ie, worse clinical outcomes) at baseline, whereas participants allocated to the usual care plus combined manual and exercise therapy presented lower WOMAC composite, pain, and function scores (ie, better clinical outcomes) (Table S3).

Moderation analysis. Results of the moderation analyses are presented in Table 3. Results of the moderation analyses for the WOMAC function and pain subscales are presented in Tables 4 and 5. BMI moderated the treatment effects of manual therapy interventions on WOMAC composite (β −4.1, 95% CI −6.6 to −1.7), function (β −3.3, 95% CI −5.1 to −1.6), and pain scores (β −0.6, 95% CI −1.3 to 0.0), where a negative β (95% CI) is the improvement in outcome associated with each one-unit increase in BMI. BMI did not moderate the effect of other interventions. Mental health moderated the effects of manual therapy on WOMAC function scores (β 16.8, 95% CI 2.2–31.4), where participants with low risk of depression presented better WOMAC function scores (Table 3). Our findings indicate that baseline levels of pain self-efficacy, quadriceps muscle strength, and education did not moderate treatment effects of interventions evaluated in the trial.

Table 3. Moderators of WOMAC composite score (1-y follow-up) changes to interventions*

Moderator and intervention	β (95% CI)	Interaction P value
BMI		
Manual therapy	−4.1 (−6.6 to −1.7)	0.001
Exercise therapy	−0.9 (−3.4 to 1.6)	0.482
Exercise + manual therapy	−1.3 (−3.9 to 1.4)	0.343
Pain self-efficacy		
Manual therapy	−2.2 (−7 to 2.7)	0.379
Exercise therapy	−4.7 (−10.5 to 1.2)	0.118
Exercise + manual therapy	1.6 (−3.9 to 7.1)	0.566
Muscle strength at baseline		
Manual therapy	−0.4 (−2.1 to 1.4)	0.681
Exercise therapy	−0.2 (−2.2 to 1.8)	0.820
Exercise + manual therapy	0.7 (−1.3 to 2.8)	0.481
Mental health		
Manual therapy	17.3 (−2.8 to 37.5)	0.092
Exercise therapy	−6.9 (−28.4 to 14.6)	0.526
Exercise + manual therapy	−15.9 (−38 to 6.2)	0.158
Education		
Manual therapy		
3 y or more of high school	−17.6 (−54.9 to 19.8)	0.354
Some tertiary training	−16.8 (−78.5 to 44.8)	0.590
Graduate degree	−0.2 (−48.3 to 47.8)	0.993
Postgraduate degree	32.8 (−32.9 to 98.5)	0.325
Exercise therapy		
3 y or more of high school	3.5 (−37.4 to 44.4)	0.865
Some tertiary training	11.1 (−42 to 64.3)	0.679
Graduate degree	5.7 (−45.4 to 56.8)	0.825
Postgraduate degree	−11.9 (−77.8 to 54)	0.722
Exercise + manual therapy		
3 y or more of high school	26.8 (−11 to 64.7)	0.163
Some tertiary training	6 (−50.5 to 62.4)	0.834
Graduate degree	47.6 (−2.6 to 97.8)	0.063
Postgraduate degree	16.3 (−40.5 to 73)	0.572

* For all analyses, usual care is the reference group. The following covariates were included in the model: age, BMI, number of years since symptom onset, quadriceps muscle strength, mental health (high risk of depression was used as the reference category), and pain self-efficacy. Regression models were adjusted for composite WOMAC scores at baseline. BMI, body mass index; CI, confidence interval; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

DISCUSSION

The aim of this study was to explore potential moderators of the effects of manual and exercise therapy on pain and function scores in individuals with hip or knee osteoarthritis by conducting a secondary analysis of the MOA trial.[4] Our findings suggest BMI moderated the effects of manual therapy interventions on WOMAC composite, function, and pain scores when compared to usual care alone. Those findings are applicable to BMI ranges observed in participants allocated to usual care and manual therapy and may not be generalizable to BMI ranges below and above the ones observed in our sample. We also found that mental health moderated treatment effects of manual therapy on WOMAC function scores, but not pain scores, and did not moderate the treatment effects of exercise therapy.

Table 4. Moderators of WOMAC function score (1-year follow-up) changes to interventions*

Moderator and intervention	β (95% CI)	Interaction *P* value
BMI		
Manual therapy	−3.3 (−5.1 to −1.6)	0.000
Exercise therapy	−1.1 (−2.9 to 0.7)	0.224
Exercise + manual therapy	−1.2 (−3.2 to 0.7)	0.199
Pain self-efficacy		
Manual therapy	−0.2 (−1.5 to 1.1)	0.748
Exercise therapy	0.0 (−1.5 to 1.4)	0.962
Exercise + manual therapy	0.3 (−1.2 to 1.9)	0.683
Muscle strength at baseline		
Manual therapy	−2.4 (−6 to 1.1)	0.179
Exercise therapy	−3.5 (−7.8 to 0.7)	0.103
Exercise + manual therapy	1.0 (−3.1 to 5)	0.642
Mental health		
Manual therapy	16.8 (2.2 to 31.4)	0.024
Exercise therapy	−3.2 (−18.7 to 12.4)	0.688
Exercise + manual therapy	−11.1 (−27.1 to 4.9)	0.172
Education		
Manual therapy		
3 y or more of high school	−14.2 (−41.8 to 13.5)	0.312
Some tertiary training	−18.9 (−64.5 to 26.8)	0.415
Graduate degree	3.9 (−31.7 to 39.4)	0.830
Postgraduate degree	18.2 (−30.4 to 66.9)	0.460
Exercise therapy		
3 y or more of high school	1.2 (−29.1 to 31.5)	0.937
Some tertiary training	0.9 (−38.4 to 40.3)	0.963
Graduate degree	9.0 (−28.8 to 46.8)	0.639
Postgraduate degree	−5.3 (−54.1 to 43.4)	0.830
Exercise + manual therapy		
3 y or more of high school	19.3 (−8.7 to 47.3)	0.175
Some tertiary training	−4.5 (−46.3 to 37.2)	0.830
Graduate degree	32.7 (−4.4 to 69.9)	0.084
Postgraduate degree	15.0 (−27.0 to 57.0)	0.480

* For all analyses, usual care is the reference group. The following covariates were included in the model: age, BMI, number of years since symptom onset, quadriceps muscle strength, mental health (high risk of depression was used as the reference category), and pain self-efficacy. Regression models were adjusted for composite WOMAC scores at baseline. BMI, body mass index; CI, confidence interval; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

Table 5. Moderators of WOMAC pain score (1-year follow-up) changes to interventions*

Moderator and intervention	β (95% CI)	Interaction *P* value
BMI		
Manual therapy	−0.6 (−1.3 to 0.0)	0.044
Exercise therapy	0.0 (−0.6 to 0.7)	0.954
Exercise + manual therapy	−0.1 (−0.7 to 0.6)	0.815
Pain self-efficacy		
Manual therapy	−0.2 (−0.7 to 0.2)	0.268
Exercise therapy	−0.2 (−0.6 to 0.3)	0.518
Exercise + manual therapy	0.2 (−0.3 to 0.8)	0.363
Muscle strength at baseline		
Manual therapy	0.4 (−0.8 to 1.6)	0.476
Exercise therapy	−1.0 (−2.4 to 0.5)	0.193
Exercise + manual therapy	0.3 (−1.1 to 1.7)	0.645
Mental health		
Manual therapy	−0.1 (−5.2 to 5.0)	0.958
Exercise therapy	−3.2 (−8.7 to 2.2)	0.247
Exercise + manual therapy	−3.6 (−9.2 to 2)	0.209
Education		
Manual therapy		
3 y or more of high school	−2.7 (−11.9 to 6.5)	0.568
Some tertiary training	−3.6 (−18.7 to 11.6)	0.644
Graduate degree	−4.6 (−16.4 to 7.3)	0.448
Postgraduate degree	8.6 (−7.6 to 24.8)	0.295
Exercise therapy		
3 y or more of high school	1.3 (−8.8 to 11.4)	0.795
Some tertiary training	5.1 (−8.0 to 18.2)	0.439
Graduate degree	−3.8 (−16.4 to 8.7)	0.547
Postgraduate degree	−6.8 (−23.0 to 9.5)	0.411
Exercise + manual therapy		
3 y or more of high school	5.7 (−3.6 to 15.1)	0.226
Some tertiary training	5.9 (−8.0 to 19.8)	0.405
Graduate degree	9.9 (−2.5 to 22.3)	0.116
Postgraduate degree	0.3 (−13.7 to 14.3)	0.965

* For all analyses, usual care is the reference group. The following covariates were included in the model: age, BMI, number of years since symptom onset, quadriceps muscle strength, mental health (high risk of depression was used as the reference category), and pain self-efficacy. Regression models were adjusted for composite WOMAC scores at baseline. BMI, body mass index; CI, confidence interval; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

Individuals with higher BMI or with lower risk of depression at baseline benefited more from manual therapy interventions than those with lower BMI or those with higher risk of depression, respectively, relative to usual care. One possible explanation is that manual therapy techniques may have provided pain relief and reduced stiffness of joints for those patients who have their joints exposed to higher compressive forces.[29] Exercise therapy may increase joint pain and negatively impact mobility in people

with higher BMI,[30] and that may be the reason for BMI not having a moderating effect on individuals who received exercise therapy. Our findings are in agreement with preliminary findings from a recent scoping review, which reported possible associations between psychological status and the effects of manual therapy on pain outcomes.[31]

Our results add to current knowledge on moderators of exercise therapy interventions. Other studies assessing pain self-efficacy as a moderator of exercise therapy for patients with knee osteoarthritis showed conflicting results.[13,14] One trial reported self-efficacy to moderate treatment effects for pain but not function scores,[13] whereas another trial found the opposite (ie, pain self-efficacy moderated treatment effects for function scores but not pain).[14] According to our findings, pain self-efficacy did not moderate the effects of interventions evaluated in the MOA trial on WOMAC composite, function, or pain scores. Based on previous and our current findings, pain self-efficacy has not demonstrated a significant and consistent effect across trials. Although in common with other posttrial moderation analyses,[13,14] our study was not powered to assess moderators of treatment effects. A recent trial showed that sleep quality moderated treatment effects of aerobic and resistance exercise, with poor sleepers better responding to that intervention.[32] Our results, combined with previous moderation analyses, add to the emerging evidence base that may help tailor future interventions.

Our study has strengths and limitations. The strengths include the following: (1) we conducted a thorough secondary analysis to explore the putative moderators of interventions tested in the MOA trial, (2) we followed best practices for analyzing and reporting moderation analyses,[33] and (3) we adjusted for potential bias by implementing inverse-probability weighting.[25] This approach allowed us to reduce the chance of bias when estimating treatment effects after excluding participants who received arthroplasty during the trial. The limitations are as follows: (1) This is a secondary analysis of the MOA trial, and we are underpowered for assessing the presence of treatment-effect modifiers.[18] This limitation is particularly relevant for education as a potential moderator because the variable includes four categories with relatively small sample sizes in each. (2) We performed multiple analyses, thus increasing risk of Type I error. However, these are exploratory analyses, with the intention of identifying possible treatment modifiers to inform the design of future interventions, so adjustment (eg, Bonferroni) was not necessary. (3) We implemented inverse-probability weighting[25] to minimize the risk of selection bias. Although this is a robust approach, there is still a chance of residual confounding on our treatment estimates.

This study explored potential moderators of treatment by conducting a secondary analysis of the MOA trial. Our exploratory analyses suggest BMI moderated the effects of manual therapy on WOMAC composite and function scores in patients with hip or knee osteoarthritis but did not moderate the effects of exercise therapy. We also found that mental health moderated the effects of manual therapy on WOMAC function scores. Based on our findings, usual care plus manual therapy may prove to be a more suitable treatment for people with higher BMI or lower risk of depression and should undergo further trials. Future trials should continue to explore which variables moderate treatment effects when testing effectiveness of nonsurgical interventions in patients with hip or knee osteoarthritis.

ACKNOWLEDGMENTS

Open access publishing facilitated by Curtin University, as part of the Wiley - Curtin University agreement via the Council of Australasian University Librarians.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Ribeiro 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. 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. doi:https://doi.org/10.1016/j.joca.2019.06.011
2. Holden MA, Hattle M, Runhaar J, et al; STEER OA Patient Advisory Group; OA Trial Bank Exercise Collaborative. Moderators of the effect of therapeutic exercise for knee and hip osteoarthritis: a systematic review and individual participant data meta-analysis. Lancet Rheumatol 2023;5(7):e386–e400. doi:https://doi.org/10.1016/S2665-9913(23)00122-4
3. Quicke JG, Runhaar J, van der Windt DA, et al. Moderators of the effects of therapeutic exercise for people with knee and hip osteoarthritis: a systematic review of sub-group analyses from randomised controlled trials. Osteoarthr Cartil Open 2020;2(4):100113. doi: https://doi.org/10.1016/j.ocarto.2020.100113
4. Abbott JH, Robertson MC, Chapple C, et al; MOA Trial team. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee: a randomized controlled trial. 1: clinical effectiveness. Osteoarthritis Cartilage 2013;21(4):525–534. doi: https://doi.org/10.1016/j.joca.2012.12.014
5. Abbott JH, Robertson MC, McKenzie JE, et al; MOA Trial team. Exercise therapy, manual therapy, or both, for osteoarthritis of the hip or knee: a factorial randomised controlled trial protocol. Trials 2009;10(1):11. doi:https://doi.org/10.1186/1745-6215-10-11
6. Abbott JH, Wilson R, Pinto D, et al; MOA Trial team. Incremental clinical effectiveness and cost effectiveness of providing supervised physiotherapy in addition to usual medical care in patients with osteoarthritis of the hip or knee: 2-year results of the MOA randomised controlled trial. Osteoarthritis Cartilage 2019;27(3):424–434. doi: https://doi.org/10.1016/j.joca.2018.12.004
7. Ribeiro DC, Lamb SE, Abbott JH. Mediators of the effects of exercise and manual therapy for people with knee and hip osteoarthritis: a secondary, exploratory analysis of the MOA trial.

Osteoarthr Cartil Open 2024;6(1):100431. doi:https://doi.org/10.1016/j.ocarto.2023.100431

8. Altman R, Alarcon G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34(5):505e14.
9. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Diagnostic and therapeutic criteria committee of the american rheumatism association. Arthritis Rheum 1986;29(8):039e49.
10. Altman RD. Criteria for classification of clinical osteoarthritis. J Rheumatol Suppl 1991;27:10e2.
11. Kraemer HC, Wilson GT, Fairburn CG, et al. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002;59(10):877–883. doi:https://doi.org/10.1001/archpsyc.59.10.877
12. Bennell KL, Dobson F, Roos EM, et al. Influence of biomechanical characteristics on pain and function outcomes from exercise in medial knee osteoarthritis and varus malalignment: exploratory analyses from a randomized controlled trial. Arthritis Care Res (Hoboken) 2015;67(9):1281–1288. doi:https://doi.org/10.1002/acr.22558
13. Lawford BJ, Hinman RS, Kasza J, et al. Moderators of effects of internet-delivered exercise and pain coping skills training for people with knee osteoarthritis: exploratory analysis of the IMPACT randomized controlled trial. J Med Internet Res 2018;20(5):e10021. doi:https://doi.org/10.2196/10021
14. Nelligan RK, Hinman RS, McManus F, et al. Moderators of the effect of a self-directed digitally delivered exercise program for people with knee osteoarthritis: exploratory analysis of a randomized controlled trial. J Med Internet Res 2021;23(10):e30768. doi:https://doi.org/10.2196/30768
15. Sandborgh M, Lindberg P, Denison E. The Pain Belief Screening Instrument (PBSI): predictive validity for disability status in persistent musculoskeletal pain. Disabil Rehabil 2008;30(15):1123–1130. doi:https://doi.org/10.1080/09638280701523200
16. Sandborgh M, Lindberg P, Denison E. Pain belief screening instrument: development and preliminary validation of a screening instrument for disabling persistent pain. J Rehabil Med 2007;39(6):461–466. doi:https://doi.org/10.2340/16501977-0072
17. Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. BMJ 2003;327(7424):1144–1146. doi:https://doi.org/10.1136/bmj.327.7424.1144
18. Brookes ST, Whitely E, Egger M, et al. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004;57(3):229–236. doi:https://doi.org/10.1016/j.jclinepi.2003.08.009
19. Pincus T, Miles C, Froud R, et al. Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. BMC Med Res Methodol 2011;11(1):14. doi:https://doi.org/10.1186/1471-2288-11-14
20. Heymans MW, Twisk JWR. Handling missing data in clinical research. J Clin Epidemiol 2022;151:185–188. doi:https://doi.org/10.1016/j.jclinepi.2022.08.016
21. StataCorp. Stata Multiple-Imputation Reference Manual: Release 11. StataCorp LP; 2009.
22. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2008. Accessed October 16, 2025. http://www.R-project.org
23. Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies. Ann Intern Med 2013;159(8):560–562. doi:https://doi.org/10.7326/0003-4819-159-8-201310150-00709
24. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017;377(14):1391–1398. doi:https://doi.org/10.1056/NEJMsm1605385
25. Murray EJ, Caniglia EC, Petito LC. Causal survival analysis: a guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence. Res Methods Med Health Sci 2021;2(1):39–49. doi:https://doi.org/10.1177/2632084320961043
26. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15(5):615–625. doi:https://doi.org/10.1097/01.ede.0000135174.63482.43
27. Grafféo N, Latouche A, Le Tourneau C, et al. ipcwswitch: an R package for inverse probability of censoring weighting with an application to switches in clinical trials. Comput Biol Med 2019;111:103339. doi:https://doi.org/10.1016/j.compbiomed.2019.103339
28. Hernan MA, Robins JM. Causal Inference: What If. CRC Press; 2023.
29. Messier SP, Pater M, Beavers DP, et al. Influences of alignment and obesity on knee joint loading in osteoarthritic gait. Osteoarthritis Cartilage 2014;22(7):912–917. doi:https://doi.org/10.1016/j.joca.2014.05.013
30. Lamb SE, Guralnik JM, Buchner DM, et al. Factors that modify the association between knee pain and mobility limitation in older women: the Women's Health and Aging Study. Ann Rheum Dis 2000;59(5):331–337. doi:https://doi.org/10.1136/ard.59.5.331
31. Keter D, Griswold D, Learman K, et al. Patient factors associated with treatment effect of manual therapy: a scoping review. JOSPT Open 2024;2(2):82–98. doi:https://doi.org/10.2519/josptopen.2024.1128
32. Hall M, Allison K, Knox G, et al. Addition of aerobic physical activity to resistance exercise for hip osteoarthritis (PHOENIX): a randomised comparative effectiveness trial. Lancet Rheumatol 2025;7(5):e343–e354. doi:https://doi.org/10.1016/S2665-9913(24)00373-4
33. Lee H, Cashin AG, Lamb SE, et al; AGReMA group. A guideline for reporting mediation analyses of randomized trials and observational studies: the AGReMA statement. JAMA 2021;326(11):1045–1056. doi:https://doi.org/10.1001/jama.2021.14075

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1247–1255
DOI 10.1002/acr.80008
Published 2026. This article is a U.S. Government work and is in the public domain in the USA. *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

New Onset of Fibromyalgia After Exposure to a Combat Environment: A Longitudinal Cohort Study

Jay B. Higgs,[1,2] Willie J. Hale,[2,3] Casey L. Straud,[2,3,4] Jim Mintz,[2] Stacey Young-McCaughan,[2,4] Kimberly D. Gomes,[2,3] Chelsea J. Sterne,[2,5] Katrina M. Lawrence-Wolff,[1] Alan J. Bartholomew,[1] Kevin M. Kelly,[6] Douglas M. Maurer,[6] Catherine Vriend,[1] Brett T. Litz,[7,8] Douglas E. Williamson,[9,10] and Alan L. Peterson,[2,3,4] for the STRONG STAR Consortium

Objective. Traumatic life events are hypothesized to be triggers for the onset of fibromyalgia. Posttraumatic stress disorder (PTSD) is a common comorbidity of fibromyalgia. However, limited prospective data are available on the development of fibromyalgia after exposure to high-magnitude stress.

Methods. This longitudinal cohort study of US military service members (N = 1,761) assessed fibromyalgia and PTSD before and upon return from combat deployment. Fibromyalgia was assessed with the 2011 questionnaire modification of the 2010 American College of Rheumatology preliminary diagnostic criteria for fibromyalgia. The PTSD Checklist Stressor-Specific Version was used to assess symptoms of PTSD.

Results. The prevalence rates of fibromyalgia in service members at predeployment (men = 2.2%; women = 2.0%) were similar to rates in civilian populations. Following deployment, the prevalence of fibromyalgia increased significantly to 8.0% in men and 11.1% in women ($P < 0.001$). The prevalence of PTSD symptoms at predeployment was 20.7% in men and 18.3% in women. The prevalence post deployment increased slightly to 22.7% in men and 25.5% in women ($P > 0.05$). By odd ratios, service members with PTSD predeployment were 2.96 times more likely to develop fibromyalgia post deployment, and those with fibromyalgia predeployment were 3.12 times more likely to develop PTSD post deployment.

Conclusions. This study provides the largest prospective data to date to support exposure to the stress of deployment to a warzone as a significant factor related to the onset of fibromyalgia. The bidirectional comorbidity between fibromyalgia and PTSD suggests a potential link in the central nervous system and has implications for management.

INTRODUCTION

Exposure to combat-related stressors during a deployment may contribute to the onset or worsening symptoms of syndromes that have been historically associated with combat exposure.[1–3] In previous conflicts, syndromes were often labeled as "unexplained symptoms," such as Gulf War syndrome, soldier's heart, shell shock, and many others.[1] High levels of stress

Presented in part at the American College of Rheumatology Annual Meeting, Washington, DC, November 11–16, 2016; a presentation was made to the Military Health System Research Symposium, Kissimmee, Florida, September 2022, and an update was presented to the same symposium on August 6, 2025.

The views expressed herein are those of the authors and do not necessarily reflect the official policy or position of the Defense Health Agency, Brooke Army Medical Center, Carl R. Darnall Army Medical Center, the Department of Defense, the Department of Veterans Affairs, or any agencies under the US Government.

Supported by the US Department of Defense (DoD) through the US Army U.S. Department of Defense, Psychological Health and Traumatic Brain Injury Program Awards, Congressionally Directed Medical Research Programs, Psychological Health and Traumatic Brain Injury Research Program (awards W81XWH-08-02-0114 to Dr Litz, W81XWH-08-2-0110 to Dr Williamson, and W81XWH-08-02-109 to Dr Peterson). Additional funding for the data analyses was provided by the U.S. Department of Defense, Traumatic Brain Injury and Psychological Health Research Program (award HT9425-24-1-1021 to Dr Peterson). Support for access to the Post-Deployment Health Assessment data was provided through the Defense Medical Surveillance System by the Armed Forces Health Surveillance Division. The funding organizations played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

[1]Brooke Army Medical Center, Joint Base San Antonio-Fort Sam Houston, San Antonio, Texas; [2]University of Texas Health Science Center at San Antonio; [3]University of Texas at San Antonio; [4]South Texas Veterans Health Care System, San Antonio; [5]The Millard and JJ Oakley School of Humanities, Education, and the Arts, Cumberland University, Lebanon, Tennessee; [6]Carl R. Darnall Army Medical Center, Fort Hood, Texas; [7]VA Boston Healthcare System, Boston, Massachusetts; [8]Boston University Chobanian & Avedisian School of Medicine, Boston, Massachusetts; [9]Duke University School of Medicine, Durham, North Carolina; [10]Durham VA Healthcare System, Durham, North Carolina.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80008.

Address correspondence via email to Jay B. Higgs, MD, at jay.higgs@gmail.com.

Submitted for publication July 21, 2025; accepted in revised form January 6, 2026.

SIGNIFICANCE & INNOVATIONS

- Understanding the pathogenesis and comorbid disorders of fibromyalgia is an essential goal in the quest for effective therapies.
- Prior studies to investigate the suspected role of prior traumatic experience have been retrospective due to the logistics of collecting data from a population before and after exposure.
- The STRONG STAR Consortium provided a unique opportunity to prospectively study a population of active duty personnel before and after deployment to a combat environment.
- Our findings of a significant increase in the prevalence of fibromyalgia and bidirectional comorbidity with posttraumatic stress disorder support a strong role for the central nervous system in the pathogenesis of fibromyalgia.

may increase the risk of the onset of new conditions, including fibromyalgia. The relationship between combat exposure and fibromyalgia was initially recognized during the Gulf War among veterans with Gulf War syndrome.[2,3] Symptom profiles of Gulf War syndrome overlapped with what was often diagnosed as fibromyalgia in civilian populations.[3]

The history of fibromyalgia diagnosis and classification is complex and controversial, leading to both patient and provider uncertainty regarding the validity of the diagnosis and treatment options.[4–8] Key features of fibromyalgia include chronic widespread pain, sleep disturbance, and cognitive dysfunction, which may overlap with other somatic and psychological disorders.[9] Among civilians, fibromyalgia has a prevalence of 2% to 8% and a negative impact on quality of life.[7,8] It costs the United States up to $35,000 per individual in annual direct health care expenditures.[10]

A common comorbidity of fibromyalgia is posttraumatic stress disorder (PTSD). PTSD diagnosis requires re-experiencing, avoidance, and hyperarousal symptoms, as well as alterations of cognition and mood after witnessing or experiencing a traumatic event.[11,12] PTSD is highly disabling, has a lifetime prevalence of 6.1% in the United States, and incurs $12,000 to $13,000 in annual direct costs per individual, with over $230 billion in direct and indirect costs in the United States.[13,14]

Cross-sectional studies have found 15.0% to 71.5% of patients with fibromyalgia also meet PTSD criteria.[15–19] In 395 patients diagnosed with fibromyalgia, investigators found 45% had PTSD versus 3% of controls.[15] Shared functional somatic symptoms support the association between PTSD and fibromyalgia.[20] When PTSD and fibromyalgia are comorbid, the severity and impact on quality of life are more significant than either alone.[18,21]

Whereas the pathophysiology of PTSD is generally accepted as involving the central nervous system, the pathophysiology of fibromyalgia remains unknown and controversial.[22] Brain imaging and functional studies suggest that central sensitization in the central nervous system may also play a primary role in fibromyalgia.[22–27] If so, fibromyalgia would be among the chronic pain conditions termed "nociplastic" by the International Association for the Study of Pain.[9] Further, the hypothalamic–pituitary–adrenal (HPA) axis is implicated in the presence of both fibromyalgia and PTSD. Patients with fibromyalgia and comorbid mood and anxiety disorders show hyperactive HPA axis activity and increased subsequent cortisol levels.[28,29] Shared pathophysiologic mechanisms in the central and peripheral nervous systems, as well as increased proinflammatory cytokine levels, also suggest a relationship between PTSD and chronic pain syndromes.[28,30]

A significant gap in the research associating fibromyalgia with PTSD is the lack of prospective studies.[18] A challenge to such research is assessing a population before and after exposure to unexpected traumatic life events. We hypothesized that military personnel who support combat deployments may be particularly susceptible to PTSD and fibromyalgia. In the present study, we investigated the conversion rates and codependency between fibromyalgia and PTSD in a large military cohort before and after deployment to a war zone, where traumatic life events are more likely to be experienced.

MATERIALS AND METHODS

Study design, population, and settings. Predeployment and postdeployment data were collected from active duty service members stationed at Fort Hood, Texas, who completed deployment to a war zone environment lasting approximately 12 months. All methods complied with the Declaration of Helsinki. The Institutional Review Boards of Brooke Army Medical Center (reviewing for Carl R. Darnall Army Medical Center; C.2011.004d) and The University of Texas Health Science Center at San Antonio (HSC20110083H) approved the study protocol. The US Army Medical Research and Development Command Office of Human Research Oversight monitored the regulatory reviews by both institutions. A total of 1,761 participants among 4,119 deployed met inclusion criteria: (1) English speaking; (2) active duty or activated reservists of any branch of the US Armed Services; (3) deployed in support of combat operations in Iraq, Afghanistan, or surrounding locations from 2010 through 2016; (4) provided informed consent; and (5) had sufficient data for the classification of fibromyalgia and PTSD both predeployment and post deployment.

Outcome measures. Diagnostic criteria for determining fibromyalgia and PTSD were consistent with criteria used in a baseline prevalence study using the same data set.[31] Probable fibromyalgia by self-report was classified using the American College of Rheumatology 2010 preliminary diagnostic criteria for

fibromyalgia, as modified for population study by questionnaire.[32,33] In brief, 19 possible body regions of pain were queried to create a widespread pain index (WPI; scale 0–19). A symptom severity score (SS; range 0–12) was calculated, comprising fatigue (0–3 points), waking unrefreshed (0–3), cognitive symptoms (0–3), headaches (0–1), abdominal pain (0–1), and depression (0–1). Study criteria for fibromyalgia were met if (1) symptoms were present for at least three months, (2) the WPI was at least 7 and the SS was at least 5 or the WPI was 3 to 6 and the SS was at least 9, and (3) the patient did not have a disorder that would otherwise sufficiently explain the pain. The last criterion was queried by questionnaire and reviewed by a rheumatologist (JBH).

PTSD was determined using the PTSD Checklist Stressor-Specific Version.[34] The frequency of occurrence of 17 symptoms over the past year was marked using 1 (not at all) to 5 (extremely) scoring. Participants were classified as having probable PTSD by self-report if they (1) reported at least one intrusion symptom, two hyperarousal symptoms, and three avoidance symptoms, each reported as present at the level of moderate or higher (scores of 3, 4, or 5) during the past month, and (2) had a total severity score of ≥50.[35–38]

Combat-related deployment stressor data for our cohort were obtained from postdeployment surveillance records in the Defense Medical Surveillance System maintained by the US Armed Forces Health Surveillance Division. The specific responses queried were as follows: (1) Did you encounter dead bodies or see people killed or wounded during this deployment? (2) Were you engaged in direct combat where you discharged a weapon? and (3) During this deployment, did you ever feel that you were in great danger of being killed?

All data were collected by questionnaire, including race and ethnicity in the form of categorical checkmark responses. The data from this study are maintained at The University of Texas Health Science Center at San Antonio in the STRONG STAR Repository. Requests for access to the data as well as for materials and the analysis code can be emailed to repository@strongstar.org.

Statistical analysis. The 1,761 participants were categorized as fibromyalgia positive (+) or negative (–) and PTSD positive (+) or negative (–) at both the predeployment and postdeployment assessments (Table 1). They were also categorized with the following four comorbidity combinations: (1) PTSD–/fibromyalgia–, (2) PTSD+/fibromyalgia–, (3) PTSD–/fibromyalgia+, and (4) PTSD+/fibromyalgia+. To determine the degree to which demographic variables were associated with four possible PTSD/fibromyalgia combinations, variables were created, and individuals were assigned to one of the four combinations. The variables were then separately cross-tabulated with each demographic variable at predeployment (Table 2) and post deployment (Table 3). Chi-square tests are reported for each demographic variable, along with standardized residuals to indicate which cells differed significantly from their expected values. Given the number of analyses conducted, the Benjamini-Hochberg correction was used to control the false discovery rate within each time point.

To examine the risk of developing PTSD or fibromyalgia at the postdeployment assessment, generalized linear mixed effects models were used with a binomial distribution and a logit link, given the dichotomous nature of the outcomes (Table 3). Separate models were examined for both PTSD and fibromyalgia. All models included fixed effects of time, group, and their interaction. The purpose of this specification was to obtain estimates of (1) the proportion of participants who were likely to have either PTSD or fibromyalgia at each time point and (2) the change in those proportions over time. Statistical tests of the significance of the latter were calculated for each demographic variable.

The final analysis determined the degree to which PTSD at predeployment was associated with fibromyalgia at post deployment and vice versa (Tables 4 and 5). Chi-square tests, standardized residuals, and odds ratios are reported for associated cross-tabulations for both the full sample (including those with a predeployment fibromyalgia and PTSD comorbidity) and restricted sample (excluding those with a predeployment fibromyalgia and PTSD comorbidity).

RESULTS

Study population. Participant demographics are presented in Table 1. A total of 1,761 service members had data sufficient for the classification of PTSD and fibromyalgia at both predeployment and post deployment. Most of the participants were men (91.3%), married (61.0%), junior noncommissioned officers (E-4 to E-6; 59.8%), and White (57.8%), with 42.2% in an ethnic minority. The average years of military service was 5.85 years, and most had completed a previous deployment (63.6%). Almost half (49.0%) deployed in a Combat Arms position.

Combat-related deployment stressor data were available for 1,630 of the 1,761 participants. Of these 1,630 individuals, 235 (14.4%) reported encountering dead bodies or seeing people killed or wounded, 33 (2.0%) reported being engaged in direct combat and having discharged a weapon, and 474 (29.0%) felt at some point they were in great danger of being killed. The total of individual persons experiencing one or more of these three stressors was 626 (38.4%).

Outcomes. Table 1 includes the proportion of individuals with either fibromyalgia or PTSD at both predeployment and post deployment, as well as the percent change from fibromyalgia– to fibromyalgia+ and from PTSD– to PTSD+ for each demographic variable. The prevalence of fibromyalgia predeployment was 2.2% for men and 2.0% for women. At the postdeployment assessment, prevalence of fibromyalgia increased by factors of

Table 1. Number and percentages of participants with FM and PTSD at predeployment and post deployment*

Demographic variable	Total		FM+ predeployment		FM+ post deployment			PTSD+ predeployment		PTSD+ post deployment		
	n	%	n	%	n	%	% Change	n	%	n	%	% Change
Gender												
Male	1,598	91.3	35	2.2	128	8.0	+5.8[a]	331	20.7	362	22.7	+2.0
Female	153	8.7	3	2.0	17	11.1	+9.2[a]	28	18.3	39	25.5	+7.2
Military pay grade												
E-1 to E-3	466	26.6	5	1.1	26	5.6	+4.5[a]	70	15.0	83	17.8	+2.8
E-4 to E-6	1,053	59.8	29	2.8	101	9.6	+6.8[a]	250	23.7	282	26.8	+3.1
E-7 to E-9	91	5.2	3	3.3	15	16.5	+13.2[b]	25	27.5	28	30.8	+3.3
O-1 to O-4	100	5.7	1	1.0	2	2.0	+1.0	4	4.0	3	3.0	−1.0
Warrant officer	20	1.1	0	0.0	1	5.0	+5.0	2	10.0	2	10.0	0.0
Deployment duty												
Combat Arms	663	49.0	16	2.4	49	7.4	+5.0[a]	157	23.7	174	26.2	+2.6
Combat Support	271	20.0	6	2.2	23	8.5	+6.3[a]	54	19.9	54	19.9	0.0
Service Support	420	31.0	15	3.6	48	11.4	+7.8[a]	92	21.9	109	26.0	+4.1
Deployment count												
None	624	36.4	3	0.5	41	6.6	+6.1[a]	82	13.1	106	17.0	+3.9
1	476	27.7	12	2.5	34	7.1	+4.6[a]	95	20.0	107	22.5	+2.5
2	365	21.3	12	3.3	31	8.5	+5.2[b]	88	24.1	92	25.2	+1.1
3	195	11.4	6	3.1	27	13.8	10.7[a]	66	33.8	63	32.3	+1.5
4+	54	3.2	5	9.3	9	16.7	+7.4	21	38.9	26	48.1	+9.2
Race and ethnicity												
White	1,011	57.8	25	2.5	86	8.5	+6.0[a]	196	19.4	215	21.3	+1.9
African American	253	14.5	3	1.2	16	6.3	+5.1[b]	60	23.7	63	24.9	+1.2
Hispanic/Latino	368	21.1	8	2.2	29	7.9	+5.7[a]	74	20.1	97	26.4	+6.3[c]
Other	116	6.6	2	1.7	14	12.1	+10.4[b]	28	24.1	26	22.4	−1.7
Marital status												
Single	277	15.8	2	0.7	11	4.0	+3.3[c]	40	14.4	47	17.0	+2.6
Dating/cohabitating	223	10.7	1	0.4	15	6.7	+6.3[a]	48	21.5	45	20.2	−1.3
Married	1,067	61.0	29	2.7	101	9.5	+6.8[a]	222	20.8	256	24.0	+3.2
Divorced	182	10.4	6	3.3	18	9.9	+6.6[c]	48	26.4	53	29.1	+2.7
Widowed	1	0.1	0	0.0	0	0.0	0.0	1	0.3	0	0.0	0.0

* Total N for the sample was 1,761; values under the "Total" category may not sum to 1,761 due to missing data for that variable. Combat Arms includes infantry, armor, field artillery, air defense artillery, special forces, and combat flight crews. Combat Support includes chemical corps, engineer, military intelligence, military police (including security forces), and signal corps. Service Support includes quartermaster, ordinance, transportation (including convoy personnel), finance, chaplain, judge advocate general, and medical (including air evacuation crews). +, positive diagnosis; E-1 to E-3, junior enlisted officers; E-4 to E-6, junior noncommissioned officers; E-7 to E-9, senior noncommissioned officers; FM, probable fibromyalgia by self-report; O, officer; PTSD, probable posttraumatic stress disorder by self-report.
[a] $P < 0.001$;
[b] $P < 0.01$;
[c] $P < 0.05$.

3.6 in men to 8.0% and 5.6 in women to 11.1% (both $P < 0.001$). The prevalence of PTSD symptoms at predeployment was 20.7% for men and 18.3% for women. The prevalence of PTSD symptoms post deployment also increased to 22.7% in men and 25.5% in women, but these increases were not statistically significant ($P > 0.05$).

From predeployment to post deployment, the change in proportion over time in service members with fibromyalgia significantly increased for all groups of enlisted service members (range of 4.5%–13.2%), those in all three deployment duty types (range of 5.0%–7.8%), and individuals with fewer than four deployments (range of 4.6%–10.7%). The proportion with fibromyalgia also increased for each race and ethnicity group (range of 5.1%–10.4%) and every marital status group (range of 3.3%–6.8%) except "widowed" (with only one member). For PTSD, only the Hispanic/Latino group showed a significant increase in PTSD over time (6.3%).

Table 2 includes the cross-tabulation of participants with various comorbid combinations of fibromyalgia and PTSD at predeployment and compares the data with that expected from random observation (standardized residuals). Nonrandom data are noted by the footnotes "[a]" (higher frequency than expected) and "[b]" (lower frequency than expected). The fibromyalgia–/PTSD+ combination accounted for most nonrandom observations. At predeployment, significant chi-square values were observed for military pay grade ($P < 0.001$), number of previous deployments ($P < 0.001$), and marital status ($P < 0.027$). The ranks of E-4 to E-6 had a higher frequency of PTSD (23.7%) relative to the standardized residuals, whereas the ranks of E-1 to E-3 (15.0%) and O-1 to O-4 (4.0%) had less. More participants who had never previously deployed were negative for both fibromyalgia and PTSD (86.9%) before deployment. More service members with three deployments had PTSD symptoms alone (31.8%), and more individuals with four or

Table 2. Cross-tabulation of participants with various comorbid FM and PTSD combinations at predeployment*

	Total		FM+/PTSD+		FM+/PTSD−		FM−/PTSD+		FM−/PTSD−				
Demographic variable	n	%	n	%	n	%	n	%	n	%	χ^2	df	*P*
Gender													
Male	1,598	91.3	24	1.5	11	0.7	307	19.2	1,256	78.6	2.02	3	0.568
Female	153	8.7	3	2.0	0	0.0	25	16.3	125	81.7			
Military pay grade													
E-1 to E-3	466	26.5	5	1.1	0	0.0	65	13.9[b]	396	85.0	46.11	15	<0.001
E-4 to E-6	1,053	59.8	19	1.8	10	0.9	231	21.9[a]	793	75.3			
E-7 to E-9	91	5.2	3	3.3	0	0.0	22	24.2	66	72.5			
O-1 to O-4	100	5.7	0	0.0	1	1.0	4	4.0[b]	95	95.0			
Warrant officer	20	1.1	0	0.0	0	0.0	2	10.0	18	90.0			
Deployment duty													
Combat Arms	663	49.0	11	1.7	5	0.8	146	22.0	501	75.6	5.33	6	0.503
Combat Support	271	20.0	3	1.1	3	1.1	51	18.8	214	79.0			
Service Support	420	31.0	12	2.9	3	0.7	80	19.0	325	77.4			
Deployment count													
None	624	36.4	3	0.5[b]	0	0.0[b]	79	12.7[b]	542	86.9[a]	74.15	12	<0.001
1	476	27.8	8	1.7	4	0.8	87	18.3	377	79.2			
2	365	21.3	8	2.2	4	1.1	80	21.9	273	74.8			
3	195	11.4	4	2.1	2	1.0	62	31.8[a]	127	65.1[b]			
4+	54	3.2	4	7.4[a]	1	1.9	17	31.5[a]	32	59.3			
Race and ethnicity													
White	1,011	57.8	17	1.7	8	0.8	179	17.7	807	79.8	12.00	9	0.213
African American	253	14.5	1	0.4	2	0.8	59	23.3	191	75.5			
Hispanic/Latino	368	21.1	8	2.2	0	0.0	66	17.9	294	79.9			
Other	116	6.6	1	0.9	1	0.9	27	23.3	87	75.0			
Marital status													
Single	277	15.8	2	0.7	0	0.0	38	13.7[b]	237	85.6	23.12	12	0.027
Dating/cohabitating	223	12.7	1	0.4	0	0.0	47	21.1	175	78.5			
Married	1,067	61.0	19	1.8	10	0.9	203	19.0	835	78.3			
Divorced	182	10.4	5	2.7	1	0.5	43	23.6	133	73.1			
Widowed	1	0.1	0	0.0	0	0.0	1	100.0	0	0.0			

* The total N for the sample was 1,761. Values under the "Total" category may not sum to 1,761 due to missing data for that variable. Combat Arms includes infantry, armor, field artillery, air defense artillery, special forces, and combat flight crews. Combat Support includes chemical corps, engineer, military intelligence, military police (including security forces), and signal corps. Service Support includes quartermaster, ordinance, transportation (including convoy personnel), finance, chaplain, judge advocate general, and medical (including air evacuation crews). +, positive diagnosis; −, negative diagnosis; E-1 to E-3, junior enlisted officers; E-4 to E-6, junior noncommissioned officers; E-7 to E-9, senior noncommissioned officers; FM, probable fibromyalgia by self-report; O, officer; PTSD, probable posttraumatic stress disorder by self-report.
[a] The marked percentage was higher than expected by chance ($P < 0.05$) according to standardized residuals. Unmarked percentages were not significantly different from that expected by chance.
[b] The marked percentage was lower than expected by chance ($P < 0.05$) according to standardized residuals. Unmarked percentages were not significantly different from that expected by chance.

more deployments met criteria for comorbid fibromyalgia and PTSD (7.4%) before deployment. Fewer single members had PTSD (13.7%) before deployment compared to partnered service members. There was no unexpected distribution by gender.

Table 3 includes the cross-tabulation of participants with combinations of fibromyalgia and PTSD at post deployment. Here, fibromyalgia played a more significant role in the nonrandom distribution of affected soldiers. There was significantly less comorbidity of fibromyalgia and PTSD in the junior enlisted ranks of E-1 to E-3 (3.4%) and officer ranks of O-1 to O-4 (1.0%), whereas the senior noncommissioned officer ranks of E-7 to E-9 experienced more comorbidity (11.0%) after deployment. At the postdeployment assessment, the rate of PTSD and fibromyalgia comorbidity was higher in service members with three (9.7%) and four or more previous deployments (14.8%) as compared to those who had not previously deployed (3.8%).

Tables 4 and 5 summarize the risk of developing new PTSD or fibromyalgia at post deployment for the full and restricted samples. Participants in the full sample included those with predeployment comorbidity (ie, were both fibromyalgia+ and PTSD+ at predeployment), whereas those in the restricted sample excluded those with predeployment comorbidity. The analysis of the risk posed by PTSD predeployment for the development of fibromyalgia post deployment (and vice versa) indicated a reciprocal comorbid relationship between these two conditions. In the full sample, those who were PTSD+ at predeployment were 2.96 (95% confidence interval [CI] 2.08–4.22) times more likely to have a new-onset diagnosis of fibromyalgia at post deployment. This odds ratio remained high at 2.43 (95% CI 1.64–3.59) in the restricted sample, which excluded baseline fibromyalgia. Similarly, those who were fibromyalgia+ in the full sample at predeployment were 3.12 (95% CI 1.63–5.95) times more likely to

Table 3. Cross-tabulation of participants with various comorbid FM and PTSD combinations at post deployment*

	Total		FM+/PTSD+		FM+/PTSD−		FM−/PTSD+		FM−/PTSD−				
Demographic variable	n	%	n	%	n	%	n	%	n	%	χ^2	df	*P*
Gender													
Male	1,598	91.3	87	5.4	41	2.6	275	17.2	1,195	74.8	1.93	3	0.586
Female	153	8.7	12	7.8	5	3.3	27	17.6	109	71.2			
Military pay grade													
E-1 to E-3	466	26.5	16	3.4[b]	10	2.1	67	14.4	373	80.0	56.51	15	<0.001
E-4 to E-6	1,053	59.8	72	6.8	29	2.8	210	19.9[a]	742	70.5			
E-7 to E-9	91	5.2	10	11.0[a]	5	5.5	18	19.8	58	63.7			
O-1 to O-4	100	5.7	1	1.0[b]	1	1.0	2	2.0[b]	96	96.0[a]			
Warrant officer	20	1.1	1	5.0	0	0.0	1	5.0	18	90.0			
Deployment duty													
Combat Arms	663	49.0	37	5.6	12	1.8	137	20.7	477	71.9	11.59	6	0.072
Combat Support	271	20.0	16	5.9	7	2.6	38	14.0	210	77.5			
Service Support	420	31.0	32	7.6	16	3.8	77	18.3	295	70.2			
Deployment count													
None	624	36.4	24	3.8[b]	17	2.7	82	13.1[b]	501	80.3	50.28	12	<0.001
1	476	27.8	22	4.6	12	2.5	85	17.9	357	75.0			
2	365	21.3	25	6.8	6	1.6	67	18.4	267	73.2			
3	195	11.4	19	9.7[a]	8	4.1	44	22.6	124	63.6			
4+	54	3.2	8	14.8[a]	1	1.9	18	33.3[a]	27	50.0[b]			
Race and ethnicity													
White	1,011	57.8	58	5.7	28	2.8	157	15.5	768	76.0	11.34	9	0.253
African American	253	14.5	11	4.3	5	2.0	52	20.6	185	73.1			
Hispanic/Latino	368	21.1	20	5.4	9	2.4	77	20.9	262	71.2			
Other	116	6.6	10	8.6	4	3.4	16	13.8	86	74.1			
Marital status													
Single	277	15.8	8	2.9	3	1.1	39	14.1	227	81.9	19.73	12	0.072
Dating/cohabitating	223	12.7	10	4.5	5	2.2	35	15.7	173	77.6			
Married	1,067	61.0	70	6.6	31	2.9	186	17.4	780	73.1			
Divorced	182	10.4	11	6.0	7	3.8	42	23.1	122	67.0			
Widowed	1	0.1	0	0.0	0	0.0	0	0.0	1	100.0			

* The total N for the sample was 1,761. Values under the "Total" category may not sum to 1,761 due to missing data for that variable. Combat Arms includes infantry, armor, field artillery, air defense artillery, special forces, and combat flight crews. Combat Support includes chemical corps, engineer, military intelligence, military police (including security forces), and signal corps. Service Support includes quartermaster, ordinance, transportation (including convoy personnel), finance, chaplain, judge advocate general, and medical (including air evacuation crews). +, positive diagnosis; −, negative diagnosis; E-1 to E-3, junior enlisted officers; E-4 to E-6, junior noncommissioned officers; E-7 to E-9, senior noncommissioned officers; FM, probable fibromyalgia by self-report; O, officer; PTSD, probable posttraumatic stress disorder by self-report.
[a] The marked percentage was higher than expected by chance ($P < 0.05$) according to standardized residuals. Unmarked percentages were not significantly different from that expected by chance.
[b] The marked percentage was lower than expected by chance ($P < 0.05$) according to standardized residuals. Unmarked percentages were not significantly different from that expected by chance.

have PTSD at post deployment. This odds ratio also remained high at 3.80 (95% CI 1.15–12.55) in the restricted sample, excluding baseline PTSD. Although a substantial portion of participants who were classified as probable fibromyalgia+ or PTSD+ at predeployment crossed the classification threshold post deployment, the odds ratios remained nearly the same in both the full and restricted samples.

DISCUSSION

This is the first investigation to prospectively assess the prevalence and comorbidity of fibromyalgia and PTSD in a large longitudinal cohort before and after deployment to a war zone. The results showed significant changes in the prevalence of fibromyalgia and PTSD from predeployment to post deployment in a young, predominantly male population who were healthy enough for a military deployment. Fibromyalgia has traditionally been thought to affect predominantly women, but at least one publication suggests that men are less likely to be given a diagnosis, thus skewing the prevalence estimates.[39] Our data showed that men and women were both affected by fibromyalgia, with predeployment to postdeployment prevalence rates rising by factors of 3.6 and 5.6, respectively.

The war zone deployment stressors experienced by our cohort were substantial, with 38% experiencing at least one of the serious threats directly related to combat. There were likely additional unassessed stressors, such as living and working in an austere combat environment, family separation, noncombat injury, or illness while deployed.

Our predeployment assessment indicated the prevalence of fibromyalgia was similar in men (2.2%) and women (2.0%). At the postdeployment assessment, fibromyalgia prevalence

Table 4. Risk of developing FM at post deployment for the full sample and restricted sample*

Sample	Outcome	Predeployment characteristic	Postdeployment characteristic		% New cases	χ^2	df	*P*	Odds ratio	95% CI LL–UL
			FM−	FM+						
Full (N = 1,761)	FM+ at post deployment	PTSD−	1,314	87[a]	6.21	39.03	1	<0.001	2.96	2.08–4.22
		PTSD+	301	59[b]	16.39					
Restricted (n = 1,723)	FM+ at post deployment	PTSD− and FM−	1,310	80	5.75	20.76	1	<0.001	2.43	1.64–3.59
		PTSD+ and FM−	290	43[b]	12.91					

* Participants in the full sample included those with predeployment comorbidity (ie, could be both FM+ and PTSD+ at predeployment), whereas those in the restricted sample excluded those with a predeployment FM. CI, confidence interval; FM, probable fibromyalgia by self-report; LL, lower limit; PTSD, probable posttraumatic stress disorder by self-report; UL, upper limit.
[a] The marked category was lower than expected by chance according to standardized residuals. Unmarked categories were not significantly different from that expected by chance.
[b] The marked category was higher than expected by chance according to standardized residuals. Unmarked categories were not significantly different from that expected by chance.

increased significantly to 8.0% in men and 11.1% in women. This provides the largest prospective data to date supporting exposure to stress in the form of deployment to a war zone as a significant factor related to the development of fibromyalgia.

PTSD prevalence increased numerically from 20.7% at predeployment to 22.7% at post deployment in men and from 18.3% to 25.5% in women. However, these increases were not statistically significant. The high baseline PTSD prevalence may be the result of previous deployments (63.6% of the sample reported at least one previous deployment), increased awareness and decreased stigmatization of PTSD symptoms in the military, or nondeployment-related traumas.[40]

This study strongly supports the bidirectional comorbidity of fibromyalgia and PTSD. Across the deployment cycle, service members who had PTSD at baseline were 2.96 times more likely to have fibromyalgia post deployment. Similarly, service members who had fibromyalgia at baseline were 3.12 times more likely to have PTSD post deployment. The odds ratios remained significant even in the restricted populations, which omitted those with PTSD and fibromyalgia comorbidity at baseline. In another study, bidirectional comorbidity between *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (ICD-10) records of fibromyalgia and ICD coding for anxiety or depression was also found among a large cohort in a Swedish health care registry.[41]

Although this comorbidity is consistent with a central nervous system component of both conditions, other explanations should be considered. These include one or more of the following: shared causation, associated risk factors, selection bias, or an unknown third associated condition.[42] Other reports suggest alternative pathophysiologic mechanisms within the central and peripheral neuroendocrine and cytokine systems, including increased levels of proinflammatory cytokines and neuropeptide Y, which influence pain signal transmission. Reduced functioning of allopregnanolone and pregnanolone has been found in both PTSD and chronic pain.[28–30]

There are limitations to the study. Participants were classified as having probable fibromyalgia or PTSD based on responses to validated but self-report questionnaires. Hence, fibromyalgia and PTSD in this study were research classifications, not clinical diagnoses.[1,39] Conversion of continuous questionnaire scales to the presence or absence of a diagnosis by a threshold score is appropriate only for study of large cohorts. Although perceived secondary gain for having a fibromyalgia and/or PTSD diagnosis is possible, this study was not connected with service members' deployment eligibility or disability rating. Differences may exist in fibromyalgia cohorts identified in studies using criteria other than that described in our methods.[5,32] Finally, our study primarily focused on exposure to potentially traumatic events in a combat zone but did not consider the degree of trauma exposure; future

Table 5. Risk of developing PTSD at post deployment for the full sample and restricted sample*

Sample	Outcome	Predeployment characteristic	Postdeployment characteristic		% new cases	χ^2	df	*P*	Odds ratio	95% CI LL–UL
			PTSD−	PTSD+						
Full (N = 1,761)	PTSD+ at post deployment	FM−	1,337	386	22.40	13.11	1	<0.001	3.12	1.63–5.95
		FM+	20	18[a]	47.37					
Restricted (n = 1,401)	PTSD+ at post deployment	FM− and PTSD−	1,140	250	17.99	5.53	1	0.019	3.80	1.15–12.55
		FM+ and PTSD−	6	5[a]	45.45					

* Participants in the full sample included those with predeployment comorbidity (ie, could be both FM+ and PTSD+ at predeployment), whereas those in the restricted sample excluded those with PTSD. CI, confidence interval; FM, probable fibromyalgia by self-report; LL, lower limit; PTSD, probable posttraumatic stress disorder by self-report; UL, upper limit.
[a] The marked category was higher than expected by chance according to standardized residuals. Unmarked categories were not significantly different from that expected by chance.

studies should examine the cumulative effect of lifetime trauma exposure on increased rates of fibromyalgia and PTSD following a combat deployment.

Our study period antedated the SARS-CoV-2 pandemic (COVID-19), eliminating long COVID and COVID-19–associated PTSD as confounding variables. Some authors have noted the similarity between long COVID and fibromyalgia.[43,44] Additional research is required to discern whether the comorbidity found in this population is relevant to the post–COVID-19 population.

In conclusion, the results of the current study found a bidirectional comorbid relationship between fibromyalgia and PTSD in a large cohort evaluated before and after deployment to a war zone. This comorbidity adds to the body of literature suggesting an important influence of fibromyalgia on the central nervous system. Clinical evaluation of PTSD and fibromyalgia should consider a possible comorbidity. Additional research is required to determine any mutual treatment benefit.[45]

ACKNOWLEDGMENTS

We thank Julie Collins for her technical manuscript review and Deanne Hargita for managing the research regulatory reviews.

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 Higgs 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. Clauw DJ, Engel CC Jr, Aronowitz R, et al. Unexplained symptoms after terrorism and war: an expert consensus statement. J Occup Environ Med 2003;45(10):1040–1048. doi:https://doi.org/10.1097/01.jom.0000091693.43121.2f
2. Jones E, Hodgins-Vermaas R, McCartney H, et al. Post-combat syndromes from the Boer war to the Gulf war: a cluster analysis of their nature and attribution. BMJ 2002;324(7333):321–324. doi:https://doi.org/10.1136/bmj.324.7333.321
3. Clauw D. The health consequences of the first Gulf war. BMJ 2003;327(7428):1357–1358. doi:https://doi.org/10.1136/bmj.327.7428.1357
4. Perrot S. Fibromyalgia: a misconnection in a multiconnected world? Eur J Pain 2019;23(5):866–873. doi:https://doi.org/10.1002/ejp.1367
5. Wolfe F, Rasker JJ. The evolution of fibromyalgia, its concepts, and criteria. Cureus 2021;13(11):e20010. doi:https://doi.org/10.7759/cureus.20010
6. Goutte J, Cathébras P. Attitudes of internal medicine specialists and rheumatologists towards chronic pain complaints. Int J Clin Pract 2021;75(10):e14463. doi:https://doi.org/10.1111/ijcp.14463
7. Clauw DJ. Fibromyalgia and related conditions. Mayo Clin Proc 2015;90(5):680–692. doi:https://doi.org/10.1016/j.mayocp.2015.03.014
8. Galvez-Sánchez CM, Duschek S, Reyes Del Paso GA. Psychological impact of fibromyalgia: current perspectives. Psychol Res Behav Manag 2019;12:117–127. doi:https://doi.org/10.2147/PRBM.S178240
9. Clauw D, Sarzi-Puttini P, Pellegrino G, et al. Is fibromyalgia an autoimmune disorder? Autoimmun Rev 2024;23(1):103424. doi:https://doi.org/10.1016/j.autrev.2023.103424
10. D'Onghia M, Ciaffi J, Ruscitti P, et al. The economic burden of fibromyalgia: a systematic literature review. Semin Arthritis Rheum 2022;56:152060. doi:https://doi.org/10.1016/j.semarthrit.2022.152060
11. Bryant RA. Post-traumatic stress disorder: a state-of-the-art review of evidence and challenges. World Psychiatry 2019;18(3):259–269. doi:https://doi.org/10.1002/wps.20656
12. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. American Psychiatric Association; 2013.
13. Goldstein RB, Smith SM, Chou SP, et al. The epidemiology of DSM-5 posttraumatic stress disorder in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions-III. Soc Psychiatry Psychiatr Epidemiol 2016;51(8):1137–1148. doi:https://doi.org/10.1007/s00127-016-1208-5
14. Davis LL, Schein J, Cloutier M, et al. The economic burden of posttraumatic stress disorder in the United States from a societal perspective. J Clin Psychiatry 2022;83(3):21m14116. doi:https://doi.org/10.4088/JCP.21m14116
15. Häuser W, Galek A, Erbslöh-Möller B, et al. Posttraumatic stress disorder in fibromyalgia syndrome: prevalence, temporal relationship between posttraumatic stress and fibromyalgia symptoms, and impact on clinical outcome. Pain 2013;154(8):1216–1223. doi:https://doi.org/10.1016/j.pain.2013.03.034
16. Gardoki-Souto I, Redolar-Ripoll D, Fontana M, et al. Prevalence and characterization of psychological trauma in patients with fibromyalgia: a cross-sectional study. Pain Res Manag 2022;2022:2114451. doi:https://doi.org/10.1155/2022/2114451
17. D'Aoust RF, Rossiter AG, Elliott A, et al. Women veterans, a population at risk for fibromyalgia: the associations between fibromyalgia, symptoms, and quality of life. Mil Med 2017;182(7):e1828–e1835. doi:https://doi.org/10.7205/MILMED-D-15-00557
18. Yavne Y, Amital D, Watad A, et al. A systematic review of precipitating physical and psychological traumatic events in the development of fibromyalgia. Semin Arthritis Rheum 2018;48(1):121–133. doi:https://doi.org/10.1016/j.semarthrit.2017.12.011
19. Coppens E, Van Wambeke P, Morlion B, et al. Prevalence and impact of childhood adversities and post-traumatic stress disorder in women with fibromyalgia and chronic widespread pain. Eur J Pain 2017;21(9):1582–1590. doi:https://doi.org/10.1002/ejp.1059
20. Nardi AE, Karam EG, Carta MG. Fibromyalgia patients should always be screened for post-traumatic stress disorder. Expert Rev Neurother 2020;20(9):891–893. doi:https://doi.org/10.1080/14737175.2020.1794824
21. Carta MG, Moro MF, Pinna FL, et al. The impact of fibromyalgia syndrome and the role of comorbidity with mood and post-traumatic stress disorder in worsening the quality of life. Int J Soc Psychiatry 2018;64(7):647–655. doi:https://doi.org/10.1177/0020764018795211
22. Gyorfi M, Rupp A, Abd-Elsayed A. Fibromyalgia pathophysiology. Biomedicines 2022;10(12):3070. doi:https://doi.org/10.3390/biomedicines10123070
23. Lee J, Lazaridou A, Paschali M, et al. A randomized controlled neuroimaging trial of cognitive behavioral therapy for fibromyalgia pain. Arthritis Rheumatol 2024;76(1):130–140. doi:https://doi.org/10.1002/art.42672

24. Xin M, Qu Y, Peng X, et al. A systematic review and meta-analysis of voxel-based morphometric studies of fibromyalgia. Front Neurosci 2023;17:1164145. doi:https://doi.org/10.3389/fnins.2023.1164145
25. Mawla I, Huang Z, Kaplan CM, et al. Large-scale momentary brain co-activation patterns are associated with hyperalgesia and mediate focal neurochemistry and cross-network functional connectivity in fibromyalgia. Pain 2023;164(12):2737–2748. doi:https://doi.org/10.1097/j.pain.0000000000002973
26. Harte SE, Harris RE, Clauw DJ. The neurobiology of central sensitization. J Appl Biobehav Res 2018;23(2):e12137. doi:https://doi.org/10.1111/jabr.12137
27. Soldatelli M, Franco ÁO, Picon F, et al. Primary somatosensory cortex and periaqueductal gray functional connectivity as a marker of the dysfunction of the descending pain modulatory system in fibromyalgia. Korean J Pain 2023;36(1):113–127. doi:https://doi.org/10.3344/kjp.22225
28. Scioli-Salter ER, Forman DE, Otis JD, et al. The shared neuroanatomy and neurobiology of comorbid chronic pain and PTSD: therapeutic implications. Clin J Pain 2015;31(4):363–374. doi:https://doi.org/10.1097/AJP.0000000000000115
29. Neeck G. Neuroendocrine and hormonal perturbations and relations to the serotonergic system in fibromyalgia patients. Scand J Rheumatol Suppl 2000;29(suppl 113):8–12. doi:https://doi.org/10.1080/030097400750001743-1
30. Baker DG, Nievergelt CM, O'Connor DT. Biomarkers of PTSD: neuropeptides and immune signaling. Neuropharmacology 2012;62(2):663–673. doi:https://doi.org/10.1016/j.neuropharm.2011.02.027
31. Lawrence-Wolff KM, Higgs JB, Young-McCaughan S, et al; STRONG STAR Consortium. Prevalence of fibromyalgia syndrome in active-duty military personnel. Arthritis Care Res (Hoboken) 2023;75(3):667–673. doi:https://doi.org/10.1002/acr.24801
32. Wolfe F, Clauw DJ, Fitzcharles MA, et al. Fibromyalgia criteria and severity scales for clinical and epidemiological studies: a modification of the ACR preliminary diagnostic criteria for fibromyalgia. J Rheumatol 2011;38(6):1113–1122. doi:https://doi.org/10.3899/jrheum.100594
33. Wolfe F, Clauw DJ, Fitzcharles M, et al. The American College of Rheumatology preliminary diagnostic criteria for fibromyalgia and measurement of symptom severity. Arthritis Care Res 2010;62(5):600–610. doi: https://doi.org/10.1002/acr.20140
34. Weathers FW, Litz BT, Herman DS, et al. The PTSD Checklist (PCL): reliability, validity, and diagnostic utility. Paper presented at: Annual Meeting of the International Society for Traumatic Stress Studies; October 1993; San Antonio, TX.
35. Bliese PD, Wright KM, Adler AB, et al. Validating the primary care posttraumatic stress disorder screen and the posttraumatic stress disorder checklist with soldiers returning from combat. J Consult Clin Psychol 2008;76(2):272–281. doi:https://doi.org/10.1037/0022-006X.76.2.272
36. Grieger TA, Cozza SJ, Ursano RJ, et al. Posttraumatic stress disorder and depression in battle-injured soldiers. Am J Psychiatry 2006;163(10):1777–1783. doi:https://doi.org/10.1176/ajp.2006.163.10.1777
37. Hoge CW, Castro CA, Messer SC, et al. Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N Engl J Med 2004;351(1):13–22. doi:https://doi.org/10.1056/NEJMoa040603
38. Smith TC, Ryan MA, Wingard DL, et al; Millennium Cohort Study Team. New onset and persistent symptoms of post-traumatic stress disorder self reported after deployment and combat exposures: prospective population based US military cohort study. BMJ 2008;336(7640):366–371. doi:https://doi.org/10.1136/bmj.39430.638241.AE
39. Vincent A, Lahr BD, Wolfe F, et al. Prevalence of fibromyalgia: a population-based study in Olmsted County, Minnesota, utilizing the Rochester Epidemiology Project. Arthritis Care Res (Hoboken) 2013;65(5):786–792. doi:https://doi.org/10.1002/acr.21896
40. Belding JN, Castañeda SF, Jacobson IG, et al; Millennium Cohort Study Team. The Millennium Cohort Study: the first 20 years of research dedicated to understanding the long-term health of US Service Members and Veterans. Ann Epidemiol 2022;67:61–72. doi:https://doi.org/10.1016/j.annepidem.2021.12.002
41. Bondesson E, Larrosa Pardo F, Stigmar K, et al. Comorbidity between pain and mental illness - evidence of a bidirectional relationship. Eur J Pain 2018;22(7):1304–1311. doi:https://doi.org/10.1002/ejp.1218
42. Valderas JM, Starfield B, Sibbald B, et al. Defining comorbidity: implications for understanding health and health services. Ann Fam Med 2009;7(4):357–363. doi:https://doi.org/10.1370/afm.983
43. Mariette X. Long COVID: a new word for naming fibromyalgia? Ann Rheum Dis 2024;83(1):12–14. doi:https://doi.org/10.1136/ard-2023-224848
44. Clauw DJ, Calabrese L. Rheumatology and Long COVID: lessons from the study of fibromyalgia. Ann Rheum Dis 2024;83(2):136–138. doi:https://doi.org/10.1136/ard-2023-224250
45. Lacefield K, Samph SP, Orbon S, et al. Integrated intervention for comorbid posttraumatic stress disorder and fibromyalgia: A pilot study of women veterans. Psychol Trauma 2020;12(7):725–729. doi:https://doi.org/10.1037/tra0000635

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1256–1264
DOI 10.1002/acr.80006
© 2026 American College of Rheumatology

Subcutaneous Versus Intravenous Tocilizumab in Aortitis Associated With Giant Cell Arteritis: Multicenter Study of 196 Patients

Carmen Secada-Gómez, MD,[1] Javier Loricera, MD, PhD,[1] Adrián Martín-Gutiérrez, MD,[1] Javier Narváez, MD, PhD,[2] Vicente Aldasoro, MD, PhD,[3] Olga Maiz, MD,[4] Paloma Vela, MD, PhD,[5] Susana Romero-Yuste, MD, PhD,[6] Eugenio de Miguel, MD, PhD,[7] Eva Galíndez-Agirregoikoa, MD,[8] Jesús C. Fernandez-López, MD,[9] Iván Ferraz-Amaro, MD, PhD,[10] Julio Sanchez-Martín, MD,[11] Patricia Moya-Alvarado, MD, PhD,[12] Cristina Campos, MD,[13] Santos Castañeda, MD, PhD,[14] and Ricardo Blanco, MD, PhD,[1] on behalf of the Tocilizumab in Giant Cell Arteritis Spanish Collaborative Group

Objective. Aortitis associated with giant cell arteritis (GCA) is a severe manifestation, potentially leading to aneurysms and aortic dissection. Tocilizumab (TCZ) has demonstrated efficacy in the treatment of GCA, both intravenously or subcutaneously administered. However, pivotal studies did not specifically evaluate aortic involvement, and no comparison of intravenous (IV) versus subcutaneous (SC) TCZ has been performed in patients with GCA-related aortitis. The objective of this study was to compare the effectiveness of TCZ according to the administration route in patients with GCA-associated aortitis under clinical practice conditions.

Methods. This was a multicenter observational study including 196 patients diagnosed with GCA-associated aortitis by imaging and treated with TCZ. Patients were grouped by administration route: IV or SC. GCA was diagnosed following the 1990 American College of Rheumatology criteria, temporal artery biopsy, and/or vascular imaging. Aortitis was identified using ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography scan. Main outcomes included EULAR remission, clinical and imaging remission, absence of systemic inflammation, and glucocorticoid-sparing effect.

Results. Of 196 patients (148 women; mean age 69.8 ± SD 9.4 years), 110 received IV TCZ and 86 SC TCZ. Baseline clinical characteristics and markers of inflammation were comparable between groups. The glucocorticoid-sparing effect was similar. At 24-month follow-up, EULAR-defined remission was significantly more frequent in the SC group (83.3% vs 80.6%; $P < 0.05$). However, rates of imaging remission and absence of systemic inflammation were comparable between treatment arms.

Conclusions. In this real-world cohort of GCA-associated aortitis, SC TCZ showed slightly greater effectiveness than IV TCZ in achieving EULAR-defined remission, whereas no significant differences were observed between both routes regarding imaging remission.

[1]Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, IDIVAL, Immunopathology Group, Santander, Cantabria, Spain; [2]Department of Rheumatology, Hospital de Bellvitge, Barcelona, Spain; [3]Department of Rheumatology, Complejo Hospitalario de Navarra, Pamplona, Spain; [4]Department of Rheumatology, Hospital Universitario de Donosti, San Sebastian, Spain; [5]Department of Rheumatology, Hospital General Universitario de Alicante, Alicante, Spain; [6]Department of Rheumatology, Complejo Hospitalario Universitario de Pontevedra, Pontevedra, Spain; [7]Department of Rheumatology, Hospital Universitario La Paz, Madrid, Spain; [8]Department of Rheumatology, Hospital Universitario de Basurto, Bilbao, Spain; [9]Department of Rheumatology, Hospital Universitario Juan Canalejo, A Coruña, Spain; [10]Department of Rheumatology, Complejo Hospitalario Universitario de Canarias, Tenerife, Spain; [11]Department of Rheumatology, Hospital Universitario 12 de Octubre, Madrid, Spain; [12]Department of Rheumatology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; [13]Department of Rheumatology, Hospital General Universitatio de Valencia, Valencia, Spain; [14]Department of Rheumatology, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria-Princesa, Madrid, Spain.

Drs Secada-Gómez and Loricera are co-first authors and contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80006.

Address correspondence via email to Ricardo Blanco, MD, PhD, at rblancovela@gmail.com.

Submitted for publication August 26, 2025; accepted in revised form December 29, 2025.

SIGNIFICANCE & INNOVATIONS

- Either subcutaneous (SC) or intravenous (IV) tocilizumab (TCZ) induces clinical and analytical remission in most patients with giant cell arteritis (GCA)–associated aortitis.
- Interestingly, SC TCZ seems to achieve higher clinical remission rates at 24 months compared to IV administration.
- Both routes of TCZ administration provide a similar glucocorticoid-sparing effect.
- Imaging response remains frequently incomplete with both administration routes, underscoring an unmet therapeutic need in GCA–aortitis.

INTRODUCTION

Giant cell arteritis (GCA) is the most common form of vasculitis in individuals over the age of 50 years, especially affecting those of Northern European ancestry.[1] Feared complications of GCA include irreversible vision loss, aortic aneurysms, and stroke.[2]

Aortitis is a frequent extracranial manifestation of GCA, observed in up to 50% of patients as detected by advanced vascular imaging such as ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) scan.[2] Although often asymptomatic, progressive inflammation may result in serious complications such as aortic aneurysms, dissection, and rupture, which carry significant risks of death.[3,4] ^{18}F-FDG PET/CT scan is considered a highly sensitive tool for detecting clinically silent aortic inflammation and identifying large-vessel involvement in GCA, even without cranial symptoms.[5–8]

Despite therapeutic advances, GCA pathogenesis remains incompletely understood. Vascular inflammation is driven by Th1 and Th17 cells, along with macrophage-derived cytokines such as interleukin-6 (IL-6), IL-1β, and tumor necrosis factor α.[9] Elevated IL-6 levels reflect systemic activity and contribute to deeper involvement of the aortic wall, particularly the media and adventitia, in large-vessel GCA. This clinically silent aortic inflammation increases the risk for aneurysm and dissection in the long term.[4–10]

High-dose glucocorticoids (GCs) remain the first-line therapy for GCA. However, prolonged use is associated with substantial toxicity and frequent relapse during GC tapering.[11] In contrast, the efficacy of conventional disease-modifying antirheumatic drugs has been inconsistent.[10]

IL-6 is a key cytokine involved in systemic inflammation and vascular injury in GCA, with elevated levels commonly observed at onset of the disease and during relapses.[12,13] Tocilizumab (TCZ), a monoclonal antibody targeting the IL-6 receptor, has demonstrated efficacy by both intravenous (IV) and subcutaneous (SC) administration. In a phase II trial, IV TCZ induced clinical remission by week 12 in 85% of patients, compared to 40% with placebo.[14] By contrast, the phase III GiACTA trial evaluated SC TCZ for sustained GC-free remission as its primary endpoint, achieving the objective at week 52 in 55% to 56% of patients versus 15% with placebo.[15]

Studies in rheumatoid arthritis have shown similar efficacy and safety between SC and IV TCZ,[16] and pharmacokinetic data confirm that SC administration reaches minimum serum concentrations comparable to monthly IV schedule.[17,18] Although TCZ was approved for GCA treatment, global regulatory authorization was primarily based on SC dosing.[15] The phase II trial exclusively investigated IV TCZ and yielded remarkable results.[14] In contrast, the pivotal GiACTA trial focused on SC administration but was not designed to assess efficacy in patients with predominant aortitis, and vascular imaging was not performed systematically.[15] Moreover, case reports describe patients with relapsing GCA recovering disease control after switching from SC to IV TCZ.[18] Based on these findings, we conducted a comparative analysis of TCZ effectiveness in GCA–aortitis according to administration route (IV versus SC).

PATIENTS AND METHODS

Patients, enrollment criteria, and study protocol. We conducted a national, multicenter, retrospective cohort study including 196 patients diagnosed with GCA-associated aortitis who were treated with TCZ between January 2014 and May 2020. All patients had previously received high-dose GCs before starting biologic therapy. To minimize selection bias, we included all patients who had received at least one dose of TCZ, regardless of clinical outcome.

The primary inclusion criterion was a diagnosis of GCA with imaging-confirmed aortitis. Patients were diagnosed in the Rheumatology or Autoimmune Disease Units of 57 Spanish referral centers. GCA diagnosis was based on the following: (a) 1990 American College of Rheumatology (ACR) criteria[19]; (b) positive temporal artery biopsy; or (c) imaging findings consistent with large-vessel vasculitis on ^{18}F-FDG PET/CT, magnetic resonance angiography (MRA), or CT angiography (CTA). Three main remission outcomes were assessed: EULAR remission, imaging remission based on PET/CT findings, and absence of systemic inflammation.[20–22]

Before initiating biologic therapy, malignancy and systemic infections, including hepatitis B and C, were ruled out. In line with national guidelines, all patients receiving biologic drugs were screened for latent tuberculosis using a tuberculin skin test (purified protein derivative), interferon–gamma release assay, and chest radiography. Prophylactic treatment for latent tuberculosis was initiated at least four weeks before TCZ and maintained for nine months.[23]

TCZ was administered following the standard regimen, either IV at 8 mg/kg every four weeks or SC at 162 mg weekly. TCZ was initiated in patients who showed inadequate response or experienced unacceptable adverse effects with conventional therapy.

The study period (2014–2020) includes years before TCZ was formally approved for GCA indication in Europe and the US in 2017. Before approval, off-label use was limited to selected patients, with most of them receiving the IV formulation due to limited availability of the SC option. The choice of IV versus SC administration was influenced by institutional logistics, patient preference, and disease severity and did vary by center. No patients in this study received both forms of dosing at different time points.

Clinical definitions and laboratory data. Clinical and laboratory data were collected at diagnosis and at months 1, 3, 6, 12, 18, and 24 of follow-up. Fever was defined as a temperature ≥38°C. Constitutional syndrome was defined as the presence of nonspecific systemic symptoms including asthenia, anorexia, and ≥5% weight loss over the preceding six months.[24,25] Arterial hypertension (AHT) was defined as systolic pressure >140 mm Hg and/or diastolic pressure ≥90 mm Hg.[26] Dyslipidemia was defined by the presence of one or more of the following data: total cholesterol levels >200 mg/dL, triglyceride levels >150 mg/dL, high-density lipoprotein cholesterol levels <40 mg/dL, or low-density lipoprotein cholesterol levels >130 mg/dL.

Routine laboratory markers included full blood cell count, erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP) serum levels, obtained at diagnosis and at each follow-up visit. ESR levels were considered elevated when at >20 mm/h in men or >25 mm/h in women. CRP levels were considered abnormal if >0.5 mg/dL, according to previously cited studies.[24,25] Anemia was defined as hemoglobin levels <12 g/dL.

Polymyalgia rheumatica (PMR) was defined according to the classification criteria proposed by the 2021 EULAR/ACR consensus.[27] EULAR remission is defined as absence of clinical signs or symptoms attributable to active GCA and normalization of acute phase reactants (CRP, ESR).[11]

Imaging remission was defined as the absence of signs of active vascular inflammation on recommended imaging modalities for GCA (^{18}F-FDG PET/CT, ultrasound, CTA, or MRA).[20] For CTA and MRA, imaging remission was defined more broadly as the absence of new or progressive vascular lesions, including wall thickening, stenosis, or aneurysm formation, compared with baseline imaging. Absence of systemic inflammation was defined as the absence of clinical signs or symptoms attributable to active GCA, normalization of acute phase reactants, and no evidence of active inflammation on imaging.[11]

Data collection. Clinical and laboratory data were retrieved according to a pre-established research protocol. Each investigator extracted information from medical records and stored it in a centralized database. Entries were independently verified by two investigators to ensure data accuracy.

Data availability statement. Data are available upon reasonable request from qualified researchers conducting rigorous, independent scientific investigations. Access will be granted following the review and approval of a research proposal and statistical analysis plan and execution of a data sharing agreement. All relevant data are included within the article.

PET/CT acquisition and image analysis. ^{18}F-FDG PET/CT was performed to confirm suspected aortitis or large-vessel vasculitis. Patients fasted at least six hours, and serum glucose was measured before injection. A dose of 7 MBq/kg of ^{18}F-FDG was administered, and images were acquired 180 minutes later using a Biograph LSO Pico 3D scanner (Siemens Medical Solutions USA, Inc, Hoffman Estates, IL). A low-dose CT was obtained for attenuation correction and anatomic localization, followed by PET acquisition at 250 seconds per bed position. Images were reconstructed using the ordered subset expectation maximization algorithm (two iterations, eight subsets).

Images were reviewed visually by two experienced nuclear medicine specialists, who assessed FDG uptake in the supra-aortic trunks, thoracic aorta, abdominal aorta, iliac arteries, and femoral and tibiofibular arteries. Remission on PET/CT was defined as vascular FDG uptake lower than liver uptake or absence of vascular signal, following Meller et al criteria.[28]

Statistical analysis. Statistical analyses were conducted using SPSS Statistics for Windows (version 29.0, IBM Corp, Armonk, NY, US). Continuous variables were tested for normality using the Kolmogorov–Smirnov test. Variables following a normal distribution were expressed as mean ± SD, whereas nonnormally distributed variables were reported as median and interquartile range (IQR). Statistical significance was set at $P < 0.05$. Group comparisons were conducted using statistical tests appropriate for the type and distribution of each variable. Continuous variables were evaluated for normality and analyzed using either *t*-tests or the Kolmogorov–Smirnov test as appropriate, whereas categorical variables were assessed with chi-square tests. Effect sizes with 95% confidence intervals were reported for all comparisons, including mean differences for continuous variables and odds ratios for categorical variables. All tests were two sided, with a significance threshold set at $P < 0.05$.

Although clinical remission, EULAR-defined remission, and absence of inflammation are time-dependent outcomes, exact dates of remission were not consistently available due to the retrospective design and predefined follow-up visits. Therefore, time-to-event methods such as Kaplan–Meier analysis or Cox regression could not be reliably applied. A complete case analysis was performed without modeling censoring or interval timing.

Ethics. The study was approved by the clinical research ethics committee (internal code 2018.080) and then by all

participant centers and was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines.

RESULTS

Baseline general features at TCZ initiation. A total of 196 patients with GCA–aortitis treated with TCZ were included, 110 (56.2%) receiving IV TCZ and 86 (43.9%) receiving SC TCZ. The mean age was 70.4 ± SD 9.5 years in the IV group and 69 ± SD 9.3 years in the SC group, with a predominance of women in both groups (73.6% IV vs 77.9% SC). The median time from GCA diagnosis to TCZ initiation was slightly shorter in the SC group (6 [IQR 2–17] months compared to 9.5 [IQR 3–21.3] months in the IV group). The proportion of patients meeting the 1990 ACR criteria was 57.3% in the IV group and 47.7% in the SC one.

Regarding clinical phenotype, primarily extracranial GCA involvement was more frequent in the SC group (41.9% vs 30.9%), whereas mixed GCA involvement was higher in the IV group (69.1% vs 58.1%). Cardiovascular risk factors, including AHT, dyslipidemia, diabetes, and smoking history, were similarly distributed between both groups. Systemic manifestations such as PMR (66.4% IV vs 67.4% SC), constitutional syndrome (43.6% IV vs 45.3% SC), and fever (15.5% IV vs 8.1% SC) showed no significant differences.

Ischemic manifestations, including headache (39.1% IV vs 36% SC), jaw claudication (14.5% IV vs 12.8% SC), visual impairment (10% IV vs 5.8% SC), and stroke (0% IV vs 1.2% SC), were also comparable between both groups. Laboratory parameters at TCZ initiation, including ESR (median 29 [IQR 14–54] IV vs median 32 [IQR 14.5–55] SC mm/1st hour) and CRP (median 1 [IQR 0–3] in both groups), did not differ significantly between both groups. Median prednisone dose was the same in both groups (15 mg/day). TCZ was used as monotherapy more frequently in SC-treated patients (77.9%) than in the IV group (62.7%). The duration of TCZ therapy differed significantly between the two groups. Patients receiving IV TCZ had a median longer treatment duration of 19 (IQR 10–27.3) months compared to 13 (IQR 6–25) months in the SC group ($P = 0.0137$). Baseline characteristics are summarized in Table 1.

GC-sparing effect. At baseline, both groups required similar doses of GCs. Over time, a gradual reduction in prednisone dosage was observed in both groups. Although the SC TCZ group showed a slightly more pronounced decrease in prednisone dose at certain time points, the overall trends between groups were comparable. Over the 24-month follow-up, prednisone tapering was similar regardless of TCZ administration route. Detailed results are displayed in Figure 1.

Main outcomes assessed (clinical remission, EULAR remission, and absence of systemic inflammation). Regarding EULAR remission, patients receiving SC TCZ achieved higher remission rates than those receiving IV TCZ across all time points evaluated. At month six, 75% of the IV group had reached EULAR remission compared to 71.4% in the SC group. At 12 months, 67.8% reached EULAR remission in the IV group vs 68.7% in the SC group. By month 24, the difference became more pronounced, 83.3% in the SC group versus 80.6% in the IV group, the latter reaching statistical significance ($P = 0.02$).

Regarding imaging remission, at month six, imaging remission was reached by 28.6% of SC TCZ patients. At month 12, the remission rates were 19.4% for the IV group and 25% for the SC group. At month 24, both groups converged, with 25% of patients in each group achieving imaging remission.

Regarding absence of systemic inflammation, although rates were higher during follow-up in the IV group at 5.5% and 8.2% at month 12 and 24, respectively, no statistical differences were found between both groups. These data are shown in Figure 2.

DISCUSSION

This multicenter retrospective study assessed long-term clinical and imaging outcomes associated with SC versus IV TCZ in patients with GCA involving the aorta. Our findings suggest that SC TCZ is associated with higher clinical remission rates at 24-month follow-up, whereas imaging responses and GC-sparing effect were comparable between both administration routes. TCZ has shown robust effectiveness in GCA, particularly in reducing flares and minimizing exposure to GCs.[29,30] Our findings suggest a potential clinical advantage of SC TCZ in achieving sustained remission in GCA-associated aortitis in clinical practice conditions. In contrast, the imaging remission rates were lower overall and did not differ significantly between both treatment groups at any time point.

However, phase II and III clinical trials did not specifically evaluate TCZ in GCA–aortitis, leaving a gap in evidence for this clinical subset. In clinical practice, complete normalization of vascular ^{18}F-FDG uptake is achieved in only one third of patients with large-vessel GCA, reflecting the more refractory nature of aortitis to imaging resolution.[31,32] This dissociation between clinical improvement and residual vascular inflammation has been previously documented, suggesting that structural changes may persist despite apparent disease control.[33,34]

Importantly, real-world data reflect variability in treatment approaches, with some centers preferring IV TCZ and others SC TCZ based on institutional logistics, patient characteristics, or disease severity.[35] Such differences in administration route selection may contribute to heterogeneity in outcomes, especially considering the pharmacokinetic divergence between both formulations and their impact on IL-6 signaling and vascular inflammation

Table 1. Baseline demographic, clinical, and imaging characteristics of patients with GCA-associated aortitis treated with TCZ*

	Overall (n = 196)	IV TCZ (n = 110)	SC TZC (n = 86)	Missing data n (%)
Baseline characteristics				
Age, mean ± SD, y	69.8 ± 9.4	70.4 ± 9.5	69 ± 9.3	0
Sex, number of female/male (% female)	148/48 (75.5)	81/29 (73.6)	67/19 (77.9)	0
Time from GCA diagnosis to TCZ onset, median [IQR], mo	7 [2–19.8]	9.5 [3–21.3]	6 [2–17]	0
Criteria, n (%)				
ACR 1990 criteria[19]	104 (53.1)	63 (57.3)	41 (47.7)	54 (27.5)
GiACTA criteria[15]	145(74)	80 (72.7)	65 (75.6)	6 (3.1)
Clinical phenotype, n (%)				
Extracranial-only	70 (35.7)	34 (30.9)	36 (41.9)	0
Mixed	126 (64.3)	76 (69.1)	50 (58.1)	0
Cardiovascular risk factors, n (%)				
High blood pressure	108 (55)	60 (54.5)	48 (55.8)	7 (3.6)
Hyperlipidemia	103 (52.6)	62 (56.4)	41 (47.7)	7 (3.6)
Diabetes mellitus	34 (17.34)	20 (18.2)	14 (16.3)	7 (3.6)
History of tobacco use	20 (10.2)	12 (10.9)	8 (9.3)	6 (3.1)
Current smoker	20 (10.2)	12 (10.9)	8 (9.3)	6 (3.1)
Systemic manifestations, n (%)				
PMR	131 (66.8)	73 (66.4)	58 (67.4)	1 (0.5)
Constitutional syndrome	87 (44.4)	48 (43.6)	39 (45.3)	1 (0.5)
Fever	24 (12.2)	17 (15.5)	7 (8.1)	1 (0.5)
Ischemic manifestations, n (%)				
Headache	74 (37.8)	43 (39.1)	31 (36)	0
Jaw claudication	27 (13.8)	16 (14.5)	11 (12.8)	8 (4.1)
Visual disturbances	16 (8.2)	11 (10)	5 (5.8)	0
Stroke	1 (0.5)	0	1 (1.2)	5 (0.5)
Laboratory parameters, median [IQR]				
ESR, mm 1st hour	32 [14–54]	29 [14–54]	32 [14.5–55]	8 (4.1)
CRP, mg/dl	1.5 [0.6–34]	1 [0–3]	1 [0–3]	5 (0.5)
Hb, g/dL	12.5 [11.3–13.8]	12.6 [11.5–13.8]	12.4 [11.05–13.9]	6 (3.1)
Imaging techniques, n (%)				
PET (number of PET/positive)	173/23 (88.3)	97/85 (88.2)	76/65 (88.4)	3 (1.5)
MRA	27 (13.8)	12 (10.9)	15 (17.4)	4 (2)
CTA	19 (9.7)	7 (6.4)	12 (14)	5 (2.6)
Temporal artery biopsy, number of biopsies/positive (%)	115/57 (58.7)	68/36 (52.9)	47/21 (44.7)	0
GC at TCZ initiation				
Patients receiving GCs, n (%)	191 (97.4)	109 (99.1)	82 (95.3)	0
Prednisone dose, median [IQR], mg/day	–	15 [10–30]	15 [9.25–30]	0
TCZ duration, median [IQR], mo	17.5 [7.25–27]	19 [10–27.3]	13 [6–25]	0
Number of TCZ monotherapy/TCZ combined (% monotherapy)	136/60 (69.4)	69/41 (62.7)	67/19 (77.9)	0

* ACR, American College of Rheumatology; CRP, C-reactive protein; CTA, computed tomography angiography; ESR, erythrocyte sedimentation rate; GC, glucocorticoid; GCA, giant cell arteritis; Hb, hemoglobin; IQR, interquartile range; IV, intravenous; MRA, magnetic resonance angiography; PET, positron emission tomography; PMR, polymyalgia rheumatica; SC, subcutaneous; TCZ, tocilizumab.

dynamics. These findings underscore the need for prospective studies specifically addressing GCA–aortitis and evaluating whether administration route influences vascular healing or imaging remission.

The observed slight superiority in clinical remission with SC TCZ may be partly explained by pharmacokinetic differences. SC administration provides more consistent serum drug concentrations with reduced peak–trough fluctuations, potentially enabling more sustained inhibition of IL-6 signaling.[36] In contrast, IV dosing produces transiently higher peaks followed by longer intervals, which may result in variable IL-6 blockade in certain patients.[37] Moreover, treatment adherence may influence therapeutic effectiveness. SC TCZ enables home-based self-administration, possibly enhancing adherence among select patient populations.[35]

Despite these clinical differences, in our series, imaging remission rates remained low and similar between groups. This observation aligns with previous reports showing that imaging remission tends to remain delayed compared to clinical improvement and that persistent vascular uptake may indicate ongoing inflammation despite apparent control of clinical and laboratory parameters.[34] Residual uptake or mural changes observed on PET/CT or vascular MRA may represent postinflammatory or fibrotic sequelae, rather than true active vasculitis. As emphasized by EULAR, imaging serves a complementary role in the management of large-vessel vasculitis but suffers from a lack of

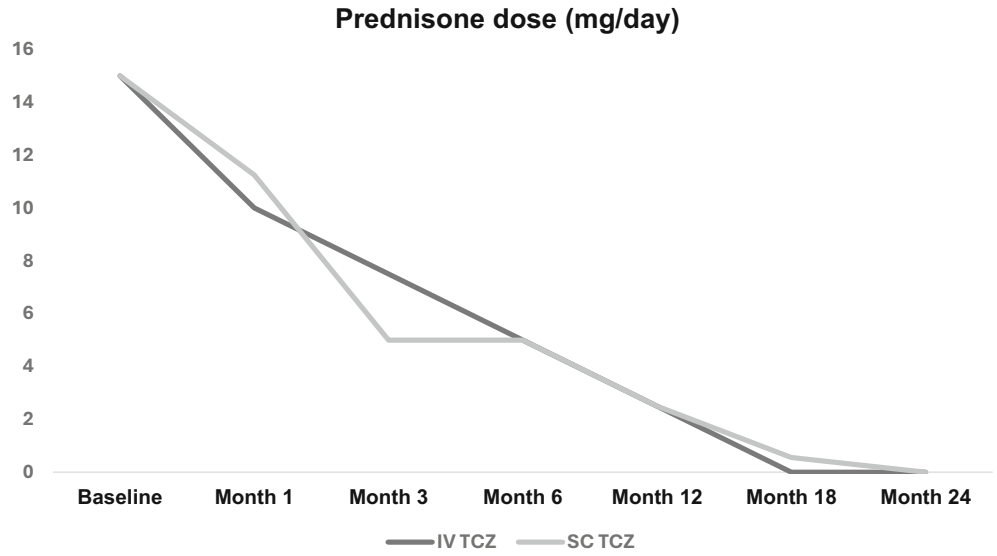

Figure 1. GC requirements over time in GCA-associated aortitis. Mean daily GC dose (prednisone-equivalent, mg/day) in patients with GCA-associated aortitis treated with either SC or IV TCZ. GC dosing was measured at predefined follow-up intervals. Data are presented as mean ± SD. GC, glucocorticoid; GCA, giant cell arteritis; IV, intravenous; SC, subcutaneous; TCZ, tocilizumab.

standardized criteria for defining remission.[20] Heterogeneity in imaging protocols and interpretation remains a recognized limitation. In addition, both treatment groups demonstrated significant reductions in cumulative GC exposure, confirming the steroid-sparing benefit of TCZ in GCA-associated aortitis.[4]

Strengths of this study include the relatively large number of patients specifically diagnosed with GCA–aortitis, the multicenter design reflecting real-world clinical conditions, and the integration of both clinical and imaging outcomes over a two-year period. In addition, the inclusion of patients regardless of clinical outcome enhances external validity by minimizing selection bias.

Nonetheless, several limitations should be considered. The retrospective design introduces risks of selection, recall, and documentation bias. Regarding methods, a limitation of our study is the inability to apply time-to-event analyses for remission outcomes. Due to the retrospective nature of data collection and lack of precise timing for outcome onset, we were unable to model censoring or use interval-based approaches. This constraint may affect the interpretation of treatment dynamics and should be considered when evaluating our findings. The use of a complete case analysis may introduce selection bias. Patients with missing data for key outcomes were excluded, potentially affecting the generalizability of our findings. Although the proportion of excluded cases was relatively low, this methodologic choice should be considered when interpreting the results.

Regarding imaging outcomes, the lack of standardized definitions for imaging response and remission in large-vessel vasculitis remains a challenge. Our study primarily relied on PET for follow-up assessment, which may limit generalizability, as PET is not universally accessible or reimbursed. PET imaging was performed using different scanner types and site-specific acquisition protocols across centers, and no centralized image reading was conducted. This heterogeneity may have introduced variability in image quality and interpretation. The definition of PET remission was intentionally stringent to ensure an objective evaluation of vascular inflammation. However, results might differ if remission were assessed using CTA or MRA, given their distinct sensitivity and specificity for detecting vascular changes. These limitations should be considered when interpreting the findings and highlight the need for further studies comparing our outcomes across different imaging modalities. Only the presence or absence of FDG uptake was evaluated in this study. Data on the degree of improvement or worsening of FDG uptake were not available, and therefore, more granular analyses using the Positron Emission Tomography Vascular Activity Score, the Time-Activity Integral, or reporting the proportion of patients with improvement or worsening could not be performed. This should be considered as a limitation when interpreting treatment efficacy based on FDG PET findings. Imaging protocols varied across centers and were interpreted locally, increasing the potential for interobserver variability.

Lastly, treatment adherence—particularly in the SC group—was not systematically assessed or documented, which may influence interpretation of therapeutic outcomes. The choice of IV versus SC formulation may introduce selection bias, as physicians could have chosen the route based on patient characteristics or logistical considerations. Due to the retrospective design, detailed information on the rationale behind formulation choice was not systematically available. Differences in treatment duration between the IV and SC groups should be considered when interpreting the results. Variability in exposure time may have influenced the observed outcomes. However, the fact that SC

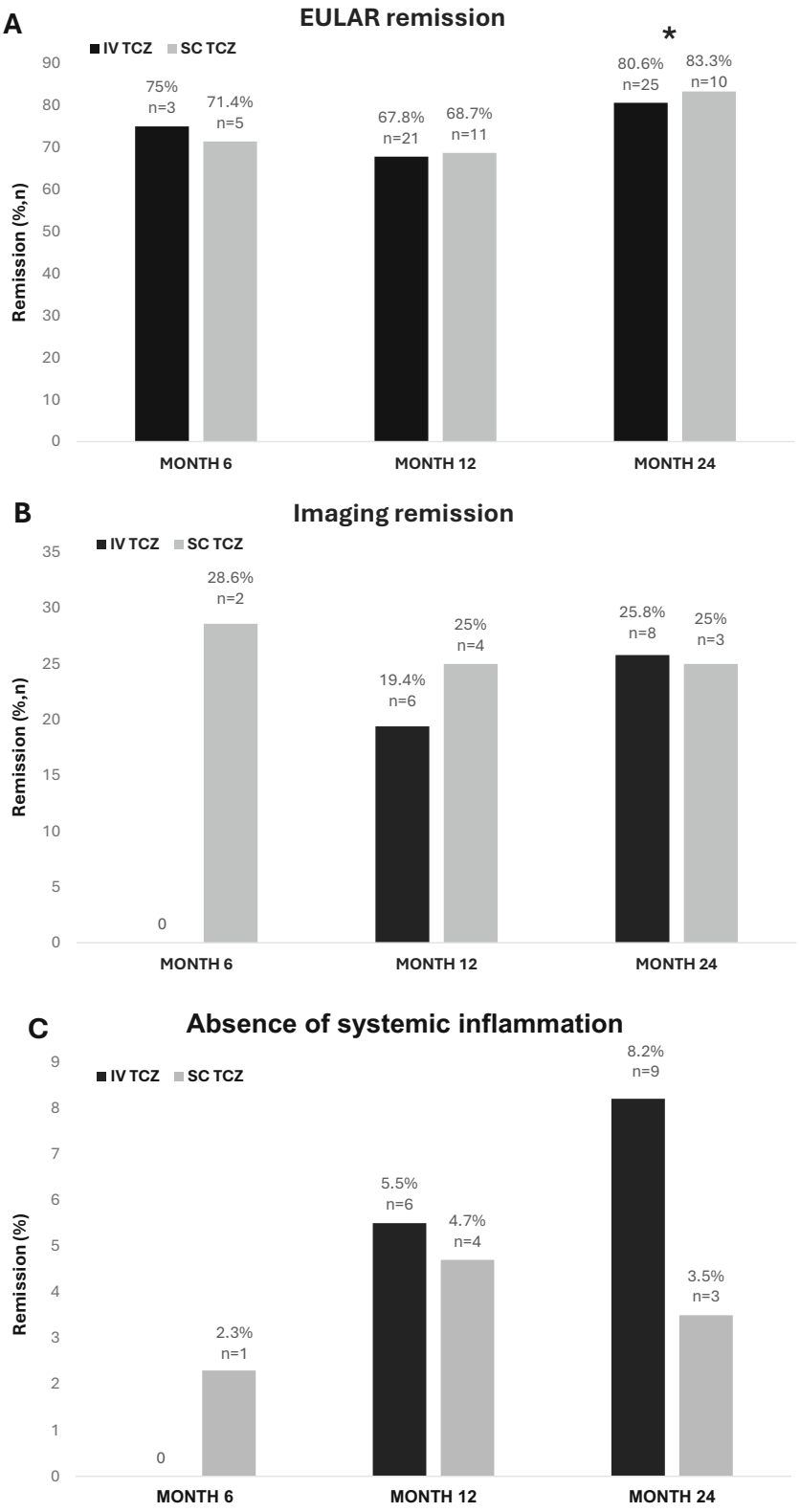

Figure 2. EULAR remission, imaging remission, and absence of systemic inflammation rates in GCA-associated aortitis treated with TCZ. (A) EULAR remission; (B) imaging remission based on PET/CT; (C) absence of systemic inflammation defined by clinical and analytical remission supported by PET/CT findings. Results represent the proportion of patients achieving remission at 6, 12, and 24 months of follow-up. Statistically significant differences between treatment groups are marked with asterisks (*$P < 0.05$). GCA, giant cell arteritis; IV, intravenous; PET/CT, positron emission tomography/computed tomography; SC, subcutaneous; TCZ, tocilizumab.

patients achieved higher remission rates despite shorter treatment duration indicates that the observed benefit is not attributable solely to exposure length. Future studies with prospective design and adjustment for treatment duration are warranted to better disentangle the effect of formulation from exposure time.

In conclusion, our findings suggest that SC TCZ may offer superior clinical effectiveness compared to IV TCZ in patients with GCA with aortitis. However, similar imaging, absence of systemic inflammation, and GC-sparing outcomes were observed, regardless of administration route. These findings may also have implications for health care resource allocation, as SC TCZ, through home-based administration, could reduce the need for infusion center visits, potentially lowering indirect costs and enhancing patient autonomy, especially in long-term management. Nonetheless, despite the practical advantages of SC TCZ, aortitis continues to represent a therapeutic challenge, as neither administration route consistently achieves vascular resolution or sustained imaging remission.

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 Blanco 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. Li KJ, Semenov D, Turk M, et al. A meta-analysis of the epidemiology of giant cell arteritis across time and space. Arthritis Res Ther 2021;23(1):82. doi:https://doi.org/10.1186/s13075-021-02450-w
2. Salvarani C, Cantini F, Boiardi L, Hunder GG. Polymyalgia rheumatica and giant-cell arteritis. N Engl J Med 2002;347(4):261–271. doi:https://doi.org/10.1056/NEJMra011913
3. Espitia O, Blonz G, Urbanski G, et al; French Study Group for Large Vessel Vasculitis (GEFA). Symptomatic aortitis at giant cell arteritis diagnosis: a prognostic factor of aortic event. Arthritis Res Ther 2021;23(1):14. doi:https://doi.org/10.1186/s13075-020-02396-5
4. Martín-Gutiérrez A, Loricera J, Narváez J, et al; Tocilizumab in Giant Cell Arteritis Spanish Collaborative Group. Effectiveness of tocilizumab in aortitis and aneurysms associated with giant cell arteritis. Eur J Intern Med 2024;129:78–86. doi:https://doi.org/10.1016/j.ejim.2024.06.013
5. Prieto-Peña D, Castañeda S, Martínez-Rodríguez I, et al. Imaging tests in the early diagnosis of giant cell arteritis. J Clin Med 2021;10(16):3704. doi:https://doi.org/10.3390/jcm10163704
6. Secada-Gómez C, Loricera J, Martín-Gutiérrez A, et al. Clinical characterization of aortitis and periaortitis: study of 134 patients from a single university hospital. Intern Emerg Med 2025;20(4):1045–1057. doi:https://doi.org/10.1007/s11739-025-03908-4
7. Loricera J, Blanco R, Hernández JL, et al. Non-infectious aortitis: a report of 32 cases from a single tertiary centre in a 4-year period and literature review. Clin Exp Rheumatol 2015;33(2 suppl 89):S19–S31.
8. Blanco R, Loricera J. 18F-FDG PET/CT in systemic autoimmune diseases and related disorders. Med Clin (Barc) 2025;164(9):487–490. doi:https://doi.org/10.1016/j.medcli.2025.01.002
9. Weyand CM, Goronzy JJ. Immunology of giant cell arteritis. Circ Res 2023;132(2):238–250. doi:https://doi.org/10.1161/CIRCRESAHA.122.322128
10. Samson M, Corbera-Bellalta M, Audia S, et al. Recent advances in our understanding of giant cell arteritis pathogenesis. Autoimmun Rev 2017;16(8):833–844. doi:https://doi.org/10.1016/j.autrev.2017.05.014
11. 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. doi:https://doi.org/10.1136/annrheumdis-2019-215672
12. Pulsatelli L, Boiardi L, Assirelli E, et al. Interleukin-6 and soluble interleukin-6 receptor are elevated in large-vessel vasculitis: a cross-sectional and longitudinal study. Clin Exp Rheumatol 2017;35 Suppl 103(1):102–110.
13. Goronzy JJ, Weyand CM. Cytokines in giant-cell arteritis. Cleve Clin J Med 2002;69(suppl 2):SII91–SII94. doi:https://doi.org/10.3949/ccjm.69.Suppl_2.SII91
14. Villiger PM, Adler S, Kuchen S, et al. Tocilizumab for induction and maintenance of remission in giant cell arteritis: a phase 2, randomised, double-blind, placebo-controlled trial. Lancet 2016;387(10031):1921–1927. doi:https://doi.org/10.1016/S0140-6736(16)00560-2
15. Stone JH, Tuckwell K, Dimonaco S, et al. Trial of tocilizumab in giant-cell arteritis. N Engl J Med 2017;377(4):317–328. doi:https://doi.org/10.1056/NEJMoa1613849
16. Burmester GR, Rubbert-Roth A, Cantagrel A, et al. A randomised, double-blind, parallel-group study of the safety and efficacy of subcutaneous tocilizumab versus intravenous tocilizumab in combination with traditional disease-modifying antirheumatic drugs in patients with moderate to severe rheumatoid arthritis (SUMMACTA study). Ann Rheum Dis 2014;73(1):69–74. doi:https://doi.org/10.1136/annrheumdis-2013-203523
17. Schmitt C, Brockwell L, Giraudon M, et al. Intravenous tocilizumab for the treatment of giant cell arteritis: a phase Ib dose-ranging pharmacokinetic bridging study. Arthritis Res Ther 2022;24(1):133. doi:https://doi.org/10.1186/s13075-022-02815-9
18. Schmalzing M, Gadeholt O, Gernert M, Tony HP, Schwaneck EC. Tocilizumab in large vessel vasculitis - different routes of administration. Open Rheumatol J 2018;12:152–159. doi:https://doi.org/10.2174/18743129018120101
19. 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. doi:https://doi.org/10.1002/art.1780330810
20. 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 2024;83(6):741–751. doi:https://doi.org/10.1136/ard-2023-224543
21. Sanchez-Alvarez C, Bond M, Soowamber M, et al. Measuring treatment outcomes and change in disease activity in giant cell arteritis: a systematic literature review informing the development of the EULAR-ACR response criteria on behalf of the EULAR-ACR response criteria in giant cell arteritis task force. RMD Open 2023;9(2):e003233. doi:https://doi.org/10.1136/rmdopen-2023-003233
22. Dejaco C, Kerschbaumer A, Aletaha D, et al. Treat-to-target recommendations in giant cell arteritis and polymyalgia rheumatica. Ann Rheum Dis 2024;83(1):48–57. doi:https://doi.org/10.1136/ard-2022-223429
23. Osorio-Chávez JS, Martínez-López D, Álvarez-Reguera C, et al. Epidemiology of latent tuberculosis in rheumatic immune-mediated inflammatory diseases—study of 1117 patients and descriptive

literature review. J Clin Med 2024;13(24):7546. doi:https://doi.org/10.3390/jcm13247546

24. Prieto-Peña D, Bernabeu P, Vela P, et al. Tocilizumab in refractory Caucasian Takayasu's arteritis: a multicenter study of 54 patients and literature review. Ther Adv Musculoskelet Dis 2021;13:1759720X211020917. doi:https://doi.org/10.1177/1759720X211020917
25. Loricera J, Castañeda S, Moriano C, et al. Tocilizumab in visual involvement of giant cell arteritis: a multicenter study of 471 patients. Ther Adv Musculoskelet Dis 2022;14:1759720X221113747. doi:https://doi.org/10.1177/1759720X221113747
26. Williams B, Mancia G, Spiering W, et al; ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J 2018;39(33):3021–3104. doi:https://doi.org/10.1093/eurheartj/ehy339
27. Ponte C, Grayson PC, Robson JC, et al. 2022 American college of rheumatology/EULAR classification criteria for giant cell arteritis. Arthritis Rheumatol 2022;74(12):1881–1889. doi:https://doi.org/10.1002/art.42325
28. Meller J, Strutz F, Siefker U, et al. Early diagnosis and follow-up of aortitis with [(18)F]FDG PET and MRI. Eur J Nucl Med Mol Imaging 2003;30(5):730–736. doi:https://doi.org/10.1007/s00259-003-1144-y
29. Blanco R, Aldasoro V, Maiz O, et al; Tocilizumab in Giant Cell Arteritis Spanish Collaborative Group. Tocilizumab in cranial and extracranial giant cell arteritis: a national multicentre study of 471 cases. Rheumatology (Oxford) 2025;64(5):2928–2936. doi:https://doi.org/10.1093/rheumatology/keae666
30. Calderón-Goercke M, Loricera J, Moriano C, et al; Tocilizumab in Giant Cell Arteritis Spanish Collaborative Group. Optimisation of tocilizumab therapy in giant cell arteritis. A multicentre real-life study of 471 patients. Clin Exp Rheumatol 2023;41(4):829–836. doi:https://doi.org/10.55563/clinexprheumatol/oqs8u9
31. Prieto Peña D, Martínez-Rodríguez I, Atienza-Mateo B, et al. Evidence for uncoupling of clinical and 18-FDG activity of PET/CT scan improvement in tocilizumab-treated patients with large-vessel giant cell arteritis. Clin Exp Rheumatol 2021;39 Suppl 129(2):69–75. doi:https://doi.org/10.55563/clinexprheumatol/mjm8fr
32. Sammel AM, Hsiao E, Schembri G, et al. Cranial and large vessel activity on positron emission tomography scan at diagnosis and 6 months in giant cell arteritis. Int J Rheum Dis 2020;23(4):582–588. doi:https://doi.org/10.1111/1756-185X.13805
33. de Boysson H, Aide N, Liozon E, et al. Repetitive 18F-FDG-PET/CT in patients with large-vessel giant-cell arteritis and controlled disease. Eur J Intern Med 2017;46:66–70. doi:https://doi.org/10.1016/j.ejim.2017.08.013
34. Grayson PC, Alehashemi S, Bagheri AA, et al. 18F-Fluorodeoxyglucose-positron emission tomography as an imaging biomarker in a prospective, longitudinal cohort of patients with large vessel vasculitis. Arthritis Rheumatol 2018;70(3):439–449. doi:https://doi.org/10.1002/art.40379
35. Alten R, Burmester GR, Breedveld FC, et al. Patient preferences and satisfaction with SC vs IV biologicals in rheumatic diseases. Adv Ther 2016;33(11):2068–2078. doi:https://doi.org/10.1007/s12325-016-0419-2
36. Zhang X, Georgy A, Rowell L. Pharmacokinetics and pharmacodynamics of tocilizumab, a humanized anti-interleukin-6 receptor monoclonal antibody, following single-dose administration by subcutaneous and intravenous routes to healthy subjects. Int J Clin Pharmacol Ther 2013;51(6):443–455. doi:https://doi.org/10.5414/CP201819
37. Bingham CO III, Ranganath VK, Halsey NA, et al. Tocilizumab: a review of pharmacokinetics and clinical efficacy. Clin Pharmacokinet 2011;50(9):563–579. doi:https://doi.org/10.2165/11590440-000000000-00000

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1265–1269
DOI 10.1002/acr.80016
Published 2026. This article is a U.S. Government work and is in the public domain in the USA. *Arthritis Care & Research* published by Wiley Periodicals LLC on behalf of American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

American College of Rheumatology
Empowering Rheumatology Professionals

BRIEF REPORT

Effect of Implementing a Dashboard With or Without a Best Practice Alert on HLA-B*58:01 Testing Rates Among Allopurinol Users at VA Medical Centers

Abimbola Fadairo-Azinge, MD, MPH,[1,2] Gabriela Schmajuk, MD, MS,[1,2] Sristi Sharma, MD, MPH,[1,2] Colleen P. Flanagan, MD,[1,2] Sara Faghihi-Kashani, MD, MPH,[1,2] Ning Zhang, MS,[1,2] and Mary A. Whooley, MD[1,2]

Objective. The American College of Rheumatology recommends HLA-B*58:01 allele testing before the initiation of allopurinol, specifically among Asian and African American/Black patients, due to their increased risk for severe hypersensitivity reactions. However, testing rates remain low at many health care facilities. This study aimed to determine whether the introduction of a clinician-facing dashboard, with or without a best practice alert (BPA), would increase HLA-B*58:01 testing rates among eligible patients at two VA medical centers.

Methods. In October 2022, we launched a clinician-facing dashboard that displayed HLA-B*58:01 testing results for all patients prescribed allopurinol at two VA medical centers. In January 2023, we added a synchronous BPA that appeared at one of the two medical centers whenever a provider entered a prescription for allopurinol. The BPA was presented as blue text on the medication order entry form and stated "HLA-B*58:01 genotyping recommended before rx for Asian or African American." Using data from the electronic health record, we compared the change in proportion of patients receiving allopurinol prescriptions who had HLA-B*58:01 genotype testing done after implementation of the dashboard (±BPA) at the two VA medical centers.

Results. From October 2022 through December 2023, the number of Asian or African American/Black patients who filled one or more prescriptions for allopurinol was 262 at the BPA + dashboard site and 330 at the dashboard only site. The percentage of Asian or African American/Black patients who had HLA-B*58:01 testing before or during the month allopurinol was prescribed increased from 8.8% in October 2022 to 35.5% in December 2023 at the dashboard + BPA site and from 2.1% to 4.5% at the dashboard only site ($P < 0.0001$ for difference in difference). The cumulative percentage of prescribers whose patient(s) completed HLA-B*58:01 testing increased from 11.7% to 46.8% at the dashboard + BPA site and from 6.9% to 10.8% at the dashboard only site ($P < 0.0001$ for difference in difference).

Conclusion. Implementing a dashboard plus simple text BPA was associated with a greater increase in the proportion of patients who completed guideline-recommend HLA-B*58:01 testing compared to implementing a dashboard alone.

INTRODUCTION

Allopurinol is a commonly prescribed medication for the treatment of gout. However, patients carrying the HLA-B*58:01 allele are at increased risk of developing severe cutaneous adverse reactions (SCARs) such as Stevens–Johnson syndrome and toxic epidermal necrolysis.[1] The mortality rate of allopurinol-induced SCARs can be as high as 25%.[2] Individuals that report Asian and African American/Black descent have a particularly high prevalence of the HLA-B*58:01 allele (7.4% and 3.8%,

Supported by awards I50-HX-003266, I50-HX-002756, and I50-HX-004148 from the Quality Enhancement Research Initiative, U.S. Department of Veterans Affairs Office of Research and Development, Health Systems Research.
[1]San Francisco VA Medical Center, San Francisco, California; [2]University of California, San Francisco.
Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.80016).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80016.
Address correspondence via email to Mary A. Whooley, MD, at Mary.whooley@va.gov.
Submitted for publication April 24, 2025; accepted in revised form January 16, 2026.

SIGNIFICANCE & INNOVATIONS

- Simple best practice alert: Including a simple text reminder on a medication order entry screen improved adherence to HLA-B*58:01 genotype testing guidelines.
- Enhanced clinical decision support: A provider-facing reminder reduced exposure to allopurinol among patients with a positive HLA-B*58:01 genotype.
- Broader clinical use: These findings emphasize the potential use of a simple medication order reminder for improving adherence to clinical guidelines, contributing to broader patient safety and quality improvement efforts in health care settings.

respectively) compared with those who identify as White (<1%).[1,3] The American College of Rheumatology (ACR) guidelines recommend HLA-B*58:01 allele testing before initiating allopurinol treatment in patients of Asian or African American/Black descent.[4]

Despite these recommendations, testing rates at major medical centers have remained low.[5] Previous studies have shown that electronic health record (EHR)–based interventions, such as dashboards and best practice alerts (BPAs), can influence clinician behavior and improve adherence to guidelines.[6] This study aimed to determine the effectiveness of implementing a dashboard alone or a dashboard combined with a BPA in increasing HLA-B*58:01 testing rates at two VA medical centers.

MATERIALS AND METHODS

The Veterans Health Administration (VHA) is the largest integrated health care system in the United States, providing comprehensive medical services to over 9 million veterans across its network of medical centers and outpatient clinics. The VHA operates as a closed health care system, meaning that patient care, including prescriptions, is primarily managed within the network. All prescriptions ordered by VA providers are processed and dispensed through VA pharmacies, ensuring a complete and centralized record of medication use.

The VA health care system uses the Computerized Patient Record System as its EHR, which allows for standardized documentation, order entry, and clinical decision support across all VA facilities. The closed nature of the system, combined with robust prescription tracking, ensures high data reliability for assessing medication prescribing patterns and adherence.

This quality improvement project was conducted at the San Francisco and Palo Alto VA Medical Centers. In October 2022, a clinician-facing dashboard was launched at two VA medical centers. This web-based dashboard can be viewed at any time and is not embedded in the EHR. The dashboard was viewable within the VA intranet. The dashboard displayed HLA-B*58:01 testing results for all patients prescribed allopurinol and compared the proportion of patients who had or had not completed testing at each facility with the national average (Supplemental Figure 1A). When the dashboard was launched, links to educational information and a recorded webinar were provided via SharePoint. Use of the dashboard was optional for providers.

At one of the medical centers, a synchronous clinical alert was implemented in January 2023. When clinicians entered an electronic order for allopurinol, they would encounter a message saying "HLA-B*58:01 genotyping recommended before rx for Asian or African American" patients (Supplemental Figure 1B). The message was truncated in this way because of limits on the number of characters. The BPA was purely informational; it did not require any acknowledgment or further action.

Data from both medical centers were collected from the VHA Clinical Data Warehouse during the 12 months before and following implementation of the BPA. Included patients were veterans alive during the study period that self-identified as of Asian or African American/Black descent and had an active prescription for allopurinol. Race and ethnicity were determined by self-selecting from a fixed set of categories. Patients were excluded if we could not verify Asian or African American/Black descent.

During each month, we calculated the proportion of patients receiving allopurinol prescriptions at each facility for whom HLA-B*58:01 genotype testing had been completed. We used difference-in-differences analysis to compare the change in proportion of patients being prescribed allopurinol who had completed HLA-B*58:01 testing between October 2022 (when the dashboard was launched) and December 2023 (12 months after the BPA was launched) at the two medical centers. We also examined the subset of patients receiving new prescriptions, defined as the absence of any previous allopurinol prescription recorded in the VA medical record. Basic characteristics of the two facilities, including facility complexity, types of prescribers, and types of patients being prescribed allopurinol, were compared using *t*-tests for continuous variables and chi-square tests for dichotomous variables. All statistical analyses were conducted using R (R Foundation, Vienna, Austria).

RESULTS

From January 2022 through December 2023, the number of Asian or African American/Black patients who filled one or more prescriptions for allopurinol was 262 at the dashboard + BPA site and 330 at the dashboard only site. The two sites were similar in complexity, patient volume, location, and academic affiliation. As compared with the dashboard alone site, the dashboard + BPA site had a higher percentage of African American/Black patients and a lower percentage of Asian patients (Table 1). Patients at the dashboard + BPA site were slightly older and more likely to have had HLA-B*58:01 testing before the study (8.8% vs 2.1%; $P = 0.0002$).

Table 1. Characteristics of patients taking allopurinol and of the facilities in which allopurinol was prescribed between 10/1/22 and 12/31/23*

	Dashboard + BPA facility (San Francisco) (n = 262)	Dashboard alone facility (Palo Alto) (n = 330)	*P* value
Patient characteristics			
Age, mean ± SD, y	67.35 ± 12.15	65.46 ± 12.69	0.09
Male sex, n (%)	259 (98.9)	329 (99.7)	0.21
Race, n (%)			0.01
Black	125 (47.7)	124 (37.6)	
Asian	137 (52.3)	206 (62.4)	
Ethnicity, n (%)			0.18
Hispanic	6 (2.3)	5 (1.5)	
Not Hispanic	252 (96.2)	312 (94.5)	
Unknown	4 (1.5)	13 (3.9)	
Duration of allopurinol treatment, n (%)			0.005
None previously	50 (19.1)	51 (15.5)	
1 to 24 mos	77 (29.4)	66 (20.0)	
>24 mos	135 (51.5)	213 (64.5)	
Most recent dose, n (%)			0.01
100–199 mg daily	105 (40.1)	101 (30.6)	
200–299 mg daily	52 (19.8)	62 (18.8)	
300–399 mg daily	84 (32.1)	148 (44.8)	
≥400 mg daily	21 (8.0)	19 (5.8)	
Facility characteristics			
University (academic) affiliation	yes	yes	
Services available	Primary, specialty, and tertiary care	Primary, specialty, and tertiary care	
Location	San Francisco Bay Area	San Francisco Bay Area	
Unique patients receiving care in 2023, n	28,905	29,094	
Unique allopurinol prescribers, n			
Rheumatologists	60	80	
Non-rheumatologists	51	22	
Total	111	102	

* BPA, best practice alert.

Between October 2022 and December 2023, the cumulative percentage of patients who had HLA-B*58:01 testing increased from 8.8% to 35.5% at the dashboard + BPA site and from 2.1% to 4.5% at the dashboard only site (Figure 1A, $P < 0.0001$). In the subset of patients who had not previously been prescribed allopurinol, the cumulative percentage who had HLA-B*58:01 testing increased from 4.0% to 44.0% at the dashboard + BPA site and from 2.0% to 13.7% at the dashboard only site (Table 2).

Between October 2022 and December 2023, the cumulative percentage of prescribers whose patient(s) completed HLA-B*58:01 testing increased from 11.7% to 46.8% at the

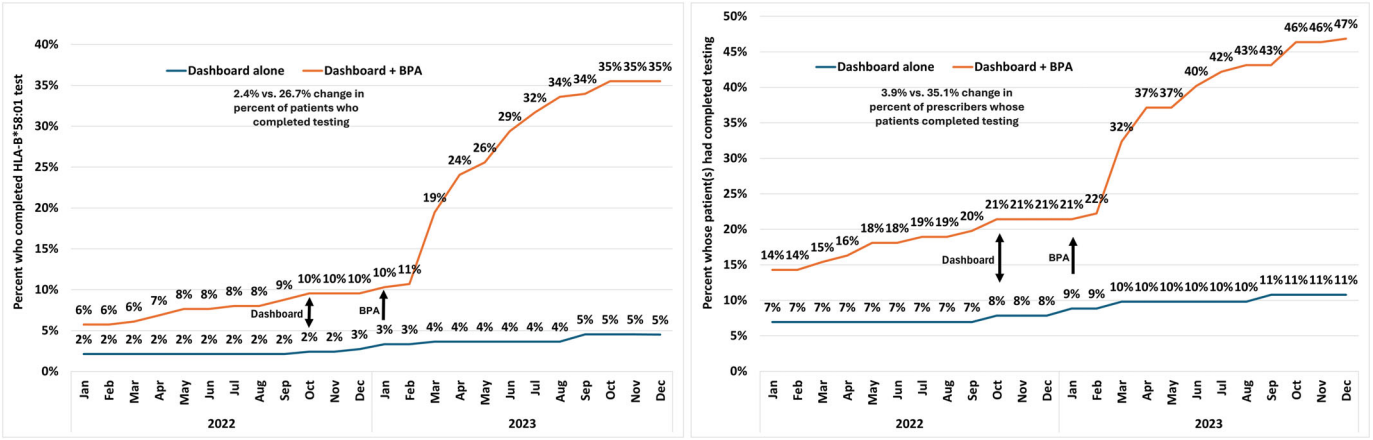

Figure 1. (Left) Cumulative percentage of Asian or African American/Black patients taking allopurinol who completed HLA-B*58:01 testing. (Right) Cumulative percentage of allopurinol prescribers whose patient(s) completed HLA-B*58:01 testing. BPA, best practice alert.

Table 2. Number of Asian or African American/Black patients taking allopurinol who completed HLA-B*58:01 test and number of prescribers whose patient(s) completed testing at dashboard + BPA versus dashboard alone facilities*

	Dashboard + BPA facility	Dashboard alone facility	*P* value
Total patients taking allopurinol, n	262	330	
Completed HLA-B*58:01 test before study, n (%)	23 (8.8)	7 (2.1)	0.0002
Completed HLA-B*58:01 test during study, n (%)	70 (26.7)	8 (2.4)	<0.0001
Total patients who completed test, n (%)	93 (35.5)	15 (4.5)	<0.0001
Patients not previously prescribed allopurinol, n	50	51	
Completed HLA-B*58:01 test before study, n (%)	2 (4.0)	1 (2.0)	0.55
Completed HLA-B*58:01 test during study, n (%)	20 (40.0)	6 (11.8)	0.001
Total patients who completed test, n (%)	22 (44.0)	7 (13.7)	0.0008
Allopurinol prescribers, n	111	102	
With at least 1 patient who completed HLA-B*58:01 test before study, n (%)	13 (11.7)	7 (6.9)	0.23
With at least 1 patient who completed HLA-B*58:01 test during study, n (%)	39 (35.1)	4 (3.9)	<0.00001
Total number of prescribers whose patient(s) completed HLA-B*58:01 test, n (%)	52 (46.8)	11 (10.8)	<0.00001

* BPA, best practice alert.

dashboard + BPA site and from 6.9% to 10.8% at the dashboard only site ($P < 0.0001$; Figure 1A). Allopurinol prescribers at the dashboard + BPA facility were more likely to order at least one HLA-B*58:01 test during the observation period than those at the dashboard only medical center (46.8% vs 10.8%, $P < 0.00001$; Figure 1B). There were five patients (four at dashboard + BPA site and one at dashboard only site) who had positive tests during the study period, and in all five patients, allopurinol was discontinued following the test.

DISCUSSION

This study found that implementation of a simple text BPA increased the proportion of Asian and African American/Black patients on allopurinol who received guideline-recommended HLA-B*58:01 testing. The improvement in testing rates was substantially greater at a BPA + dashboard medical center than at a dashboard only medical center. In addition, the simple text BPA increased the proportion of allopurinol prescribers who ordered HLA-B*58:01 genotyping. Taken together, these results demonstrate that a simple text BPA visible on a medication order entry screen provided a timely reminder to clinicians without requiring any extra clicks and emphasized the importance of following ACR guidelines.[4]

The increased testing rates observed at the BPA medical center suggest that synchronous tools that provide immediate prompts may be more effective in changing provider behavior compared to tools that require clinicians to seek out information independently. Further, the significant improvement associated with the roll-out of the clinical alert suggests that this effect outweighs that of an asynchronous dashboard. These findings align with prior studies that have shown EHR-based clinical alerts to be effective in improving adherence to evidence-based practices when delivered at the point of care.[7] Health systems that are investing in clinical decision support tools may find these alerts more impactful than an asynchronous tool. Alternatively, increased interoperability within EHR systems may unlock greater potential from asynchronous dashboards.

The implications of this study are particularly relevant for health systems seeking to optimize the use of clinical decision support tools. Future research should focus on evaluating the long-term outcomes of implementing such interventions, including the impact on patient safety and adverse event reduction, as well as exploring the potential for similar approaches in other areas of clinical practice. Additionally, we hope to expand this alert to other VHA sites to further its positive effect on veterans.

Some limitations of this study include the self-identification of race, which may have led to inappropriate exclusion of "mixed race" individuals for whom we could not verify Asian or African American/Black descent. Also, although the two sites had similar characteristics, variations in provider practices between the locations may have contributed to differences in outcomes. For example, greater engagement with the dashboard (at the dashboard + BPA site than at the dashboard only site) could have contributed to the increased screening rates at the dashboard + BPA site. Because this was an observational study, we cannot exclude the possibility that greater engagement with the dashboard resulted in higher rates of testing.

In summary, the combination of a clinician-facing dashboard and a synchronous BPA was associated with a significant increase in HLA-B*58:01 testing rates among high-risk patients prescribed allopurinol compared to the dashboard alone. These findings suggest that, BPAs, when used in conjunction with dashboards, can effectively influence clinician behavior and improve adherence to guideline-recommended testing, ultimately enhancing patient safety. These findings have implications beyond HLA-B*58:01 testing and could possibly be used in broader clinical settings across other specialties.

REFERENCES

1. Wu R, Cheng YJ, Zhu LL, et al. Impact of HLA-B*58:01 allele and allopurinol-induced cutaneous adverse drug reactions: evidence from 21 pharmacogenetic studies. Oncotarget 2016;7(49):81870–81879. doi:https://doi.org/10.18632/oncotarget.13250
2. Yaseen W, Auguste B, Zipursky J. Allopurinol hypersensitivity syndrome. CMAJ 2023;195(13):E483. doi:https://doi.org/10.1503/cmaj.221575
3. Quach C, Galen BT. HLA-B*5801 testing to prevent allopurinol hypersensitivity syndrome. JAMA Intern Med 2018;178(9):1260–1261. doi:https://doi.org/10.1001/jamainternmed.2018.3556
4. FitzGerald JD, Dalbeth N, Mikuls T, et al. 2020 American College of Rheumatology guideline for the management of gout. Arthritis Care Res. 2020;72(6):744–760. doi:https://doi.org/10.1002/acr.241
5. Sullivan J, Ware A, Tarasovsky G, et al. Facility variation in HLA-B*58:01 Allele testing for Asian and Black patients receiving allopurinol in the Veterans Affairs Healthcare System [abstract]. Arthritis Rheumatol 2023;75(suppl 9). https://acrabstracts.org/abstract/facility-variation-in-hla-b5801-allele-testing-for-asian-and-black-patients-receiving-allopurinol-in-the-veterans-affairs-healthcare-system/
6. Khairat SS, Dukkipati A, Lauria HA, et al. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018;5(2):e22. doi:https://doi.org/10.2196/humanfactors.9328
7. Coma E, Medina M, Méndez L, et al. Effectiveness of electronic point-of-care reminders versus monthly feedback to improve adherence to 10 clinical recommendations in primary care: a cluster randomized clinical trial. BMC Med Inform Decis Mak 2019;19(1):245. doi:https://doi.org/10.1186/s12911-019-0976-8

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1270–1276
DOI 10.1002/acr.80023
© 2026 American College of Rheumatology.

BRIEF REPORT

Variations in the Adult Rheumatology Fellowship Training in the United States: Opportunities and Challenges

Debendra Pattanaik, MD, Roxanna Sabghi, MD, Govind S. Bindra, MD, and Syed Raza, MD

Objective. Adult rheumatology fellowship training programs in the United States exhibit considerable curricular diversity while staying within the framework provided by American College of Rheumatology (ACR) and Accreditation Council for Graduate Medical Education (ACGME). This study aims to characterize the variation in key training components across programs, providing a descriptive overview of current practices.

Methods. We developed a set of survey questions using Qualtrics to capture curricular elements among ACGME-accredited adult rheumatology fellowship programs in the United States. One current trainee from each program was invited to participate.

Results. We received responses from 91 out of 126 programs available at the time of study, leading to a response rate of 72.2%. Although most programs incorporate core elements recommended by ACGME and ACR, substantial variability was observed in areas such as ambulatory and inpatient training, elective offerings, procedural experience, didactic structure, interdisciplinary exposure, mentorship, and research opportunities.

Conclusion. Our findings highlight the diverse approaches taken by rheumatology fellowship programs in structuring their curricula. Rather than advocating for uniformity, this variability may reflect the adaptability of programs to local resources and institutional strengths. Further exploration of how these differences influence educational outcomes could inform future efforts to enhance rheumatology training.

INTRODUCTION

The field of rheumatology has undergone substantial advances in the understanding of disease pathogenesis and therapeutic development, contributing to growing interest in the specialty among internal medicine trainees.[1,2] In the United States, both the number of rheumatology fellowship positions and applicants has increased steadily from 2015 through 2025, according to the National Resident Matching Program.[3] Currently, there are 132 rheumatology fellowship programs that operate nationwide.[4]

In July 2024, the Accreditation Council for Graduate Medical Education (ACGME) updated program requirements for rheumatology fellowship training, emphasizing program personnel, educational curriculum, evaluation processes, and the learning and working environment.[5] The American College of Rheumatology (ACR) had previously updated its core curriculum in 2023 to align with ACGME expectations.[6] This curriculum outlines recommended educational activities to support clinical competency, professional development, and scholarly engagement, including mentored clinical experience, didactics, committee participation, research projects, and presentations.[6] Although ACGME and ACR recommendations provide detailed guidance for program directors in developing fellowship training curricula, there remains no standardized rheumatology fellowship curriculum in the United States. This flexibility allows programs to tailor training to local resources and priorities but may also result in heterogeneity in training experiences nationwide. This lack of standardization provided the rationale for our national survey examining variability in fellowship training practices. A prior study in pediatric rheumatology has demonstrated substantial variability in fellowship training experiences and underscored the need for more uniform curricular frameworks and objective measures of trainee evaluation.[7] Comparable data in adult rheumatology fellowship training across programs are limited.

The primary objective of this study was to characterize the structure of adult rheumatology fellowship programs, including

University of Tennessee Health Science Center, Memphis.

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.80023).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80023.

Address correspondence via email to Debendra Pattanaik, MD, at dpattana@uthsc.edu.

Submitted for publication May 26, 2025; accepted in revised form February 4, 2026.

SIGNIFICANCE & INNOVATIONS

- There is currently no standardized curriculum for adult rheumatology fellowship training in the United States. Although the Accreditation Council for Graduate Medical Education and American College of Rheumatology offer guidelines, individual programs retain flexibility in designing their own curricula.
- To date, no study has described how these curricula vary across programs.
- Our study is the first to outline key similarities and differences in curricular components among adult rheumatology fellowship programs nationwide.
- By highlighting areas of variability, this study may inform future discussions on curriculum development and provide impetus for further research into the impact of this diversity on trainee readiness for independent practice.

clinical rotations, didactics, electives, procedural training, and research opportunities. We sought to identify similarities and differences across programs and to highlight areas that may inform future curriculum development and target educational interventions by individual programs.

MATERIALS AND METHODS

We conducted a national, cross-sectional survey of adult rheumatology fellowship programs in the United States to characterize current training structures, clinical experiences, and educational practices. We developed a set of survey questions aimed at identifying similarities and differences among rheumatology fellowship programs, and they are included in the supplementary material section for reference (Supplementary File 1). We formulated survey questions along the lines of a similar survey conducted for pediatric rheumatology fellowships[7] in 2012. A focus group, consisting of four other faculty members and three fellows of our own program who were not study authors, reviewed the questions and provided feedback. We also sent the survey questionnaires to another fellowship program for review and feedback. The questionnaire and study were approved by the institutional review board (IRB) of the University of Tennessee Health Science Center (IRB 19-06841-XP).

This was a program-based survey in which one current fellow from each ACGME-accredited adult rheumatology fellowship program was invited to participate; individual fellows were not randomly sampled. Participation was voluntary and anonymous. We obtained the list of current ACGME-accredited programs, which is located on the website of the American Medical Association.[4] Trainees and their contact information were identified through a detailed search of the rheumatology provider directory published by the ACR,[8] as well as the websites of individual fellowship programs. If contact information was unavailable, program coordinators and program directors from individual programs were contacted for assistance. Selected fellows were contacted by email between January and June 2023 to participate in the online survey using Qualtrics. A $50 Amazon gift card was sent to fellows who completed the survey. Statistical analyses were performed using Stata version 17 MP (StataCorp, College Station, TX). Continuous variables were summarized as medians with interquartile ranges and compared across groups using the Kruskal-Wallis equality-of-populations rank test. Categorical variables were expressed as frequencies and percentages and compared using the Pearson chi-square test. All tests were two sided, and a P value <0.05 was considered statistically significant.

RESULTS

Demographics and fellowship structure. We received responses from 91 of 126 programs, yielding a response rate of 72.2%. A detailed summary of responses is presented here. Among respondents, 58% (53 of 91) identified as women and 42% (38 of 91) as men. Most fellows were in their second year (55%), followed by first year (41%), and only 4% were in their third year. Fellowship duration was two years (79%), with 3% lasting three years and 18% offering either the option of two or three years. Programs reported a mean of 2.23 (SD 0.89, range 1–5) first-year fellows and 2.11 (SD 0.75, range 2–5) second-year fellows. Most programs had no third-year fellows (mean 0.31, range 0–3). The number of core faculty regularly involved in fellowship training varied: 37% reported 4 to 6 faculty, 27% reported 7 to 9 faculty, 26% reported ≥10 faculty, and 9% reported 1 to 3 faculty (Figure 1).

Ambulatory experience. There was substantial variability in weekly clinic half-days. During the first year, clinic time ranged from one to nine half-days per week (mean 4.18, SD 1.49) and in the second year from zero to nine half-days (mean 3.71, SD 1.74). On average, fellows saw 2.02 new (SD 1.15) and 4.99 follow-up patients (SD 1.52) per clinic half-day. Most fellows (81%) reported serving medically indigent or underserved populations in their continuity clinic. At least one continuity clinic was located at a Veterans Affairs Medical Center for 57% of fellows. Fellows' participation in disease-specific clinics varied, with lupus clinic being most common (42.9%), followed by scleroderma (20.9%) and osteoporosis and metabolic bone disease clinics (16.5%).

Inpatient experience. The inpatient rheumatology service functioned primarily as a consult service for most fellows (95%; Figure 2). Only 1.1% reported a purely primary service, whereas 4.4% reported a combined consult and primary model. Coverage typically involved one fellow (79%), with fewer programs assigning two (13%) or three to four fellows (8%) concurrently. Fellows

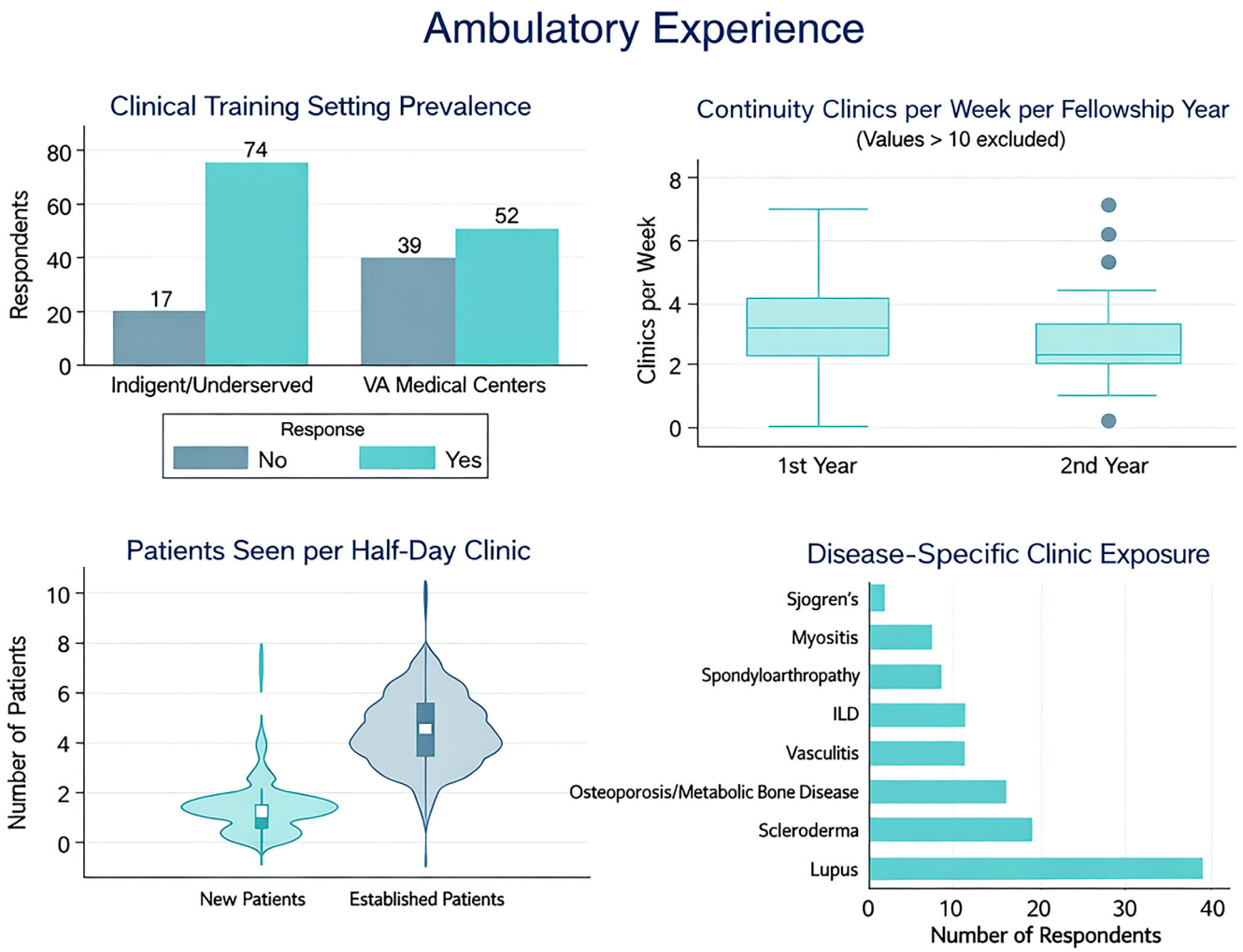

Figure 1. Ambulatory experience.

reported a wide range of total consult months during training. Total inpatient consult months ranged from 3 to 26 months (mean 9.59, SD 4.39). The most reported inpatient volume was 6 to 10 consults per week (36%), followed by 11 to 15 patients (22%), 0 to 5 patients (20%), 16 to 20 patients (14%), and >20 patients (8%). Additionally, fellows reported the number of weekend days on call per month, counting each Saturday and Sunday separately. Weekend call ranged from one to six days per month (mean 2.6, SD 1.2).

Perception of the program. Given the outpatient focus of rheumatology practice, fellows were asked to assess the balance of their training. Most trainees (63%) felt their program was well-balanced, whereas 28% perceived an outpatient-heavy experience and 9% an inpatient-heavy experience.

Electives and procedural training. Outpatient pediatric exposure ranged from 0 to 24 months (mean 1.96, SD 3.59), whereas inpatient pediatric exposure was limited (mean 0.49, SD 1.77) (Figure 3). Common elective rotations included musculoskeletal radiology (59%), followed by dermatology (40%) and orthopedic surgery (35%).

A formal injection clinic was available in 58% of programs. Most fellows (90%) felt "very experienced" with knee injections. Shoulder injections were the next most performed, with 52% feeling "very experienced" and 45% reporting "some experience." In contrast, the hip was the least frequently injected joint, with 73% indicating "little or no experience." Formal musculoskeletal ultrasonography training was reported by 55% of fellows. Use of polarized microscopy for crystal identification was limited: 25% reported regular use, 52% reported availability but rare use, and 23% reported no availability.

Academic activities. Most of the fellows (91%) spent two to five hours per week on formal didactics, with three hours most reported (34%), followed by four hours (27%), two hours (18%), and five hours (12%; Supplementary Figure 1). Interdisciplinary conferences or clinics were most frequently reported with

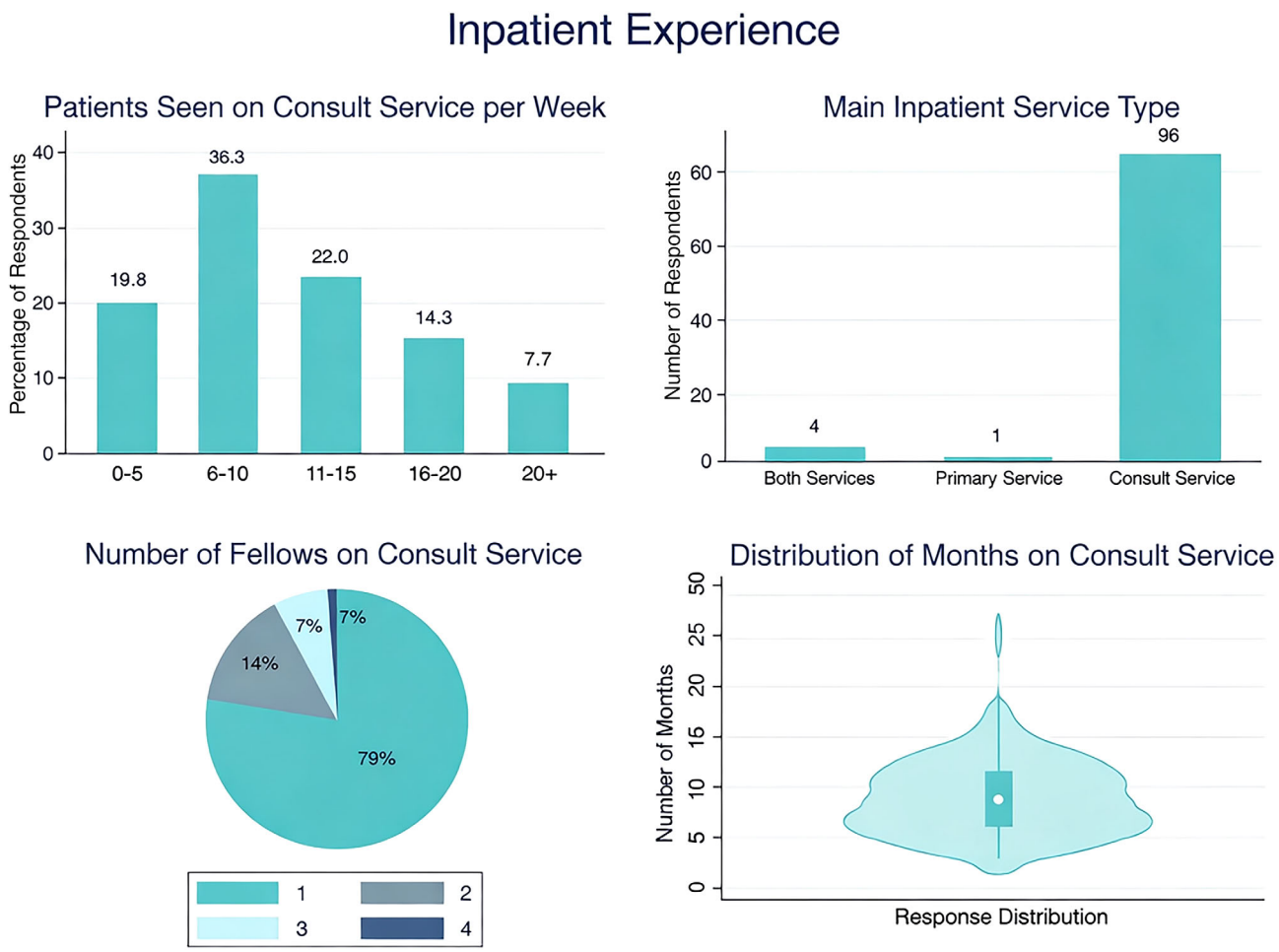

Figure 2. Inpatient experience. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80023/abstract.

radiology (54 respondents), followed by dermatology (36 respondents), nephrology (35 respondents), and pulmonary (32 respondents). Fewer than 10 fellows reported collaborations with neurology or ophthalmology. Only one fellow reported collaboration with orthopedic surgery. Most fellows were engaged in clinical research (67 of 91) and quality improvement initiatives (64 of 91). Basic science research was less common, with only 11 fellows participating in such projects. Scholarly output included peer-reviewed presentations (41%), regional presentations (48%), national presentations (45%), and international presentations (8%). Formal mentorship programs were unavailable to most fellows (59%), although 41% reported access to structured mentorship.

Responses based on fellowship training year. The responses were similar between first-year and second/third-year fellows (Supplementary Tables S1–S4). Established patient volume per clinic per half-day was higher among second/third-year fellows compared to first-year fellows (P = 0.017). Procedural experience, particularly with small joint and wrist injections, was significantly greater among second/third-year fellows (P = 0.010 and P = 0.033, respectively). A trend toward fewer clinic half-days in later years was also observed (P = 0.063).

DISCUSSION

There is wide variation in clinical practice among rheumatologists across the United States.[9,10] Multiple factors contribute to this variability,[11] despite the availability of evidence-based clinical practice guidelines published by professional societies such as ACR, which are intended to promote consistent, high-quality care across disease areas.[12] Differences in fellowship training experiences may play a key role in shaping variability in clinical decision-making and practice patterns among rheumatologists.[2] Similar patterns have been observed in other specialties, in which trainees' clinical practice patterns are shaped by their residency or fellowship experiences.[13] Such variation in training across different training programs is not inherently problematic and may reflect the unique strengths and weaknesses of individual programs.

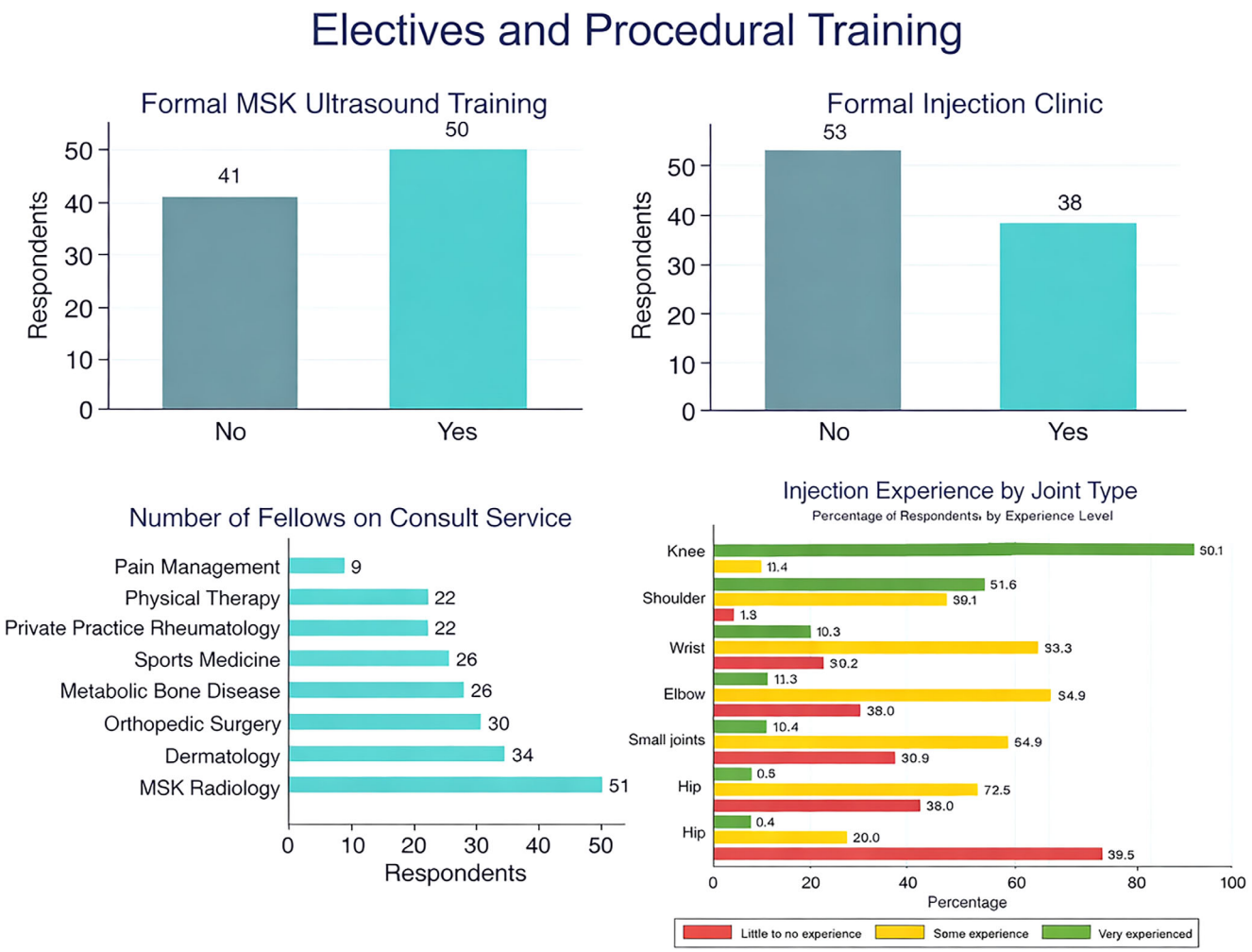

Figure 3. Electives and procedural training. MSK, musculoskeletal. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80023/abstract.

In this study, we surveyed 91 of 126 rheumatology fellowship programs, achieving a response rate of 72.2%, which we consider robust. We found substantial variability across multiple domains of fellowship training, although most programs appeared to meet key ACGME and ACR requirements. Responses from first-year and second/third-year fellows were similar, with expected differences reflecting progression of clinical and procedural experiences as they advanced through the fellowship training. Below, we examine specific components of fellowship training in greater detail.

The 2023 ACR curricular update does not specify recommendations regarding ambulatory clinic structure.[6] In contrast, the 2006 ACR curriculum did recommend a full-day clinic during the first year and a half-day clinic during the second year.[6] Our survey demonstrated wide variation in the ambulatory clinic experience across programs. Fellows reported an average of four half-day clinics per week during their first year and three half-day clinics during the second year. Patient volume and clinic structure varied but were consistent with ACGME expectations.[5] These differences raise important questions regarding how ambulatory exposure influences learning outcomes and readiness for independent practice.

We observed substantial variation in the scope and structure of inpatient rheumatology training. Although some programs provide extensive inpatient consult exposure, others emphasize outpatient care. The impact of these differing training models on fellows' preparedness for various practice environments warrants further study.

The 2006 ACR fellowship curriculum emphasized multidisciplinary collaboration through combined conferences or clinics, recognizing the complex and systemic nature of many rheumatologic diseases.[6] Suggested specialties included dermatology, orthopedics, rehabilitative medicine, ophthalmology, and pediatric rheumatology. The 2023 ACR update did not revise this component.[6]

In our survey, most fellows reported participation in multidisciplinary activities, most commonly with radiology, dermatology, nephrology, pulmonology, ophthalmology, and orthopedics.

Although these interactions were widely reported, their structure and frequency varied considerably. Further investigation is needed to determine how multidisciplinary engagement influences clinical competency and team-based care.

Half of the fellows reported participation in a radiology elective, whereas approximately one-third of fellows completed electives in dermatology and orthopedics. This variability reflects differences in institutional resources and fellows' interests. Although elective flexibility allows programs to tailor fellows' training, it also raises questions about which electives most effectively support clinical practice and how fellows prioritize available options.

A major update in the 2023 ACR rheumatology curriculum formally integrated pediatric rheumatology into the core curriculum with defined knowledge expectations for adult trainees.[6] Our survey revealed wide variability in pediatric exposure, with outpatient experience ranging from 0 to 24 months (mean 1.96 months) and minimal inpatient exposure (mean 0.49 months), a pattern that may affect fellows' confidence and underscores the need to better define optimal pediatric training.

Although most fellows reported proficiency in knee injections, experience with other joints, including shoulders and hips, was less common, suggesting a limited procedural skill set among many trainees. We assessed the presence of formal joint injection clinics but not ultrasonography-guided injection training specifically. Slightly more than half of fellows reported access to structured musculoskeletal ultrasonography education. Although polarized microscopy was widely available, routine clinical use was reported by only 23% of programs. These findings underscore substantial variability in procedural training and how procedural competence is achieved.

Training in rheumatology requires exposure to a broad spectrum of patient populations, including those who are medically underserved or have unique health care needs. These populations often present complex social and medical challenges, offering fellows valuable experience in delivering equitable and comprehensive care. In our survey, 81% of programs reported that their fellows care for medically underserved populations, and 57% indicated that their trainees regularly treat patients within a Veterans Affairs Medical Center. These experiences provide fellows with critical insights into health care disparities and help cultivate skills in managing complex cases within constrained systems. Many programs also reported disease-specific clinics, most commonly lupus clinics, followed by scleroderma and metabolic bone disease clinics, offering focused experience in chronic rheumatic conditions.

All participating programs reported offering formal didactic education, typically two to five hours per week, using a variety of educational formats. A total of 70% of fellows reported involvement in clinical research or quality improvement projects, reflecting strong engagement in clinical scholarship but limited exposure to laboratory-based research due to differences in program-level research infrastructure. Overall, our findings parallel those reported in pediatric rheumatology fellowship training,[7] in which programs met accreditation and professional society standards yet demonstrated substantial variability in clinical exposure, procedural training, and mentorship. Although ACGME and ACR guidelines define core training components, they do not ensure uniform educational experiences across programs. This variability appears inherent to guideline-based curricula rather than a deficiency in accreditation standards and highlights the value of periodic national assessments to better characterize training practices and identify opportunities for improvement.

Several limitations warrant consideration. As with other cross-sectional surveys, selection bias is possible because participation depends on the identification and availability of a single fellow per program and may overrepresent programs with greater transparency in curricular structure or stronger educational cultures. Self-reported data introduce potential recall and reporting bias. Some of the responses (eg, core faculty numbers) could have reflected individual trainee perspectives rather than program-level measures, and inclusion of fellows at different training stages may have influenced reported exposure.

This national survey of adult rheumatology fellowship programs demonstrates substantial variability in training experiences across the United States. Although such variability reflects programmatic diversity and flexibility, it also raises questions regarding consistency in educational exposure, preparedness for independent practice, and potential implications for patient care. Future research will be helpful to evaluate how different training structures influence trainee outcomes, with the goal of informing best practices while preserving programmatic autonomy.

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 Pattanaik 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. Burmester GR, Bijlsma JWJ, Cutolo M, et al. Managing rheumatic and musculoskeletal diseases - past, present and future. Nat Rev Rheumatol 2017;13(7):443–448. doi:10.1038/nrrheum.2017.95
2. Miloslavsky EM, Bolster MB. Addressing the rheumatology workforce shortage: A multifaceted approach. Semin Arthritis Rheum 2020; 50(4):791–796. doi:10.1016/j.semarthrit.2020.05.009
3. National Resident Matching Program. Results and Data: Specialties Matching Service, 2025 Appointment Year. February 13, 2025. Accessed April 27, 2025. https://www.nrmp.org/match-data/2025/02/results-and-data-specialties-matching-service-2025-appointment-year/

4. AAMC. Rheumatology (Internal Medicine). 2025. Accessed April 27, 2025. https://systems.aamc.org/eras/erasstats/par/display.cfm?NAV_ROW=PAR&SPEC_CD=150
5. ACGME. ACGME Program Requirements for Graduate Medical Education in Rheumatology. 2024. Accessed April 27, 2025. https://www.acgme.org/globalassets/pfassets/programrequirements/2025-reformatted-requirements/150_rheumatology_2025_reformatted.pdf
6. American College of Rheumatology. Core Curriculum Outline for Rheumatology Fellowship Programs. Updated 2023. Accessed April 27, 2025. https://assets.contentstack.io/v3/assets/bltee37abb6b278ab2c/blt5451c0437b4e2037/core-curriculum-outline.pdf
7. Patwardhan A, Henrickson M, Laskosz L, et al. Current pediatric rheumatology fellowship training in the United States: what fellows actually do. Pediatr Rheumatol Online J 2014;12(1):8. doi:10.1186/1546-0096-12-8
8. Rheumatology provider directory. American College of Rheumatology. Accessed April 27, 2025. https://my.rheumatology.org/rheumatology-provider-directory
9. Hochberg MC, Perlmutter DL, Hudson JI, et al. Preferences in the management of osteoarthritis of the hip and knee: results of a survey of community-based rheumatologists in the United States. Arthritis Care Res 1996;9(3):170–176. doi:10.1002/1529-0131(199606)9:3<170:AID-ANR1790090304>3.0.CO;2-K
10. Henke CJ, Epstein WV. Practice variation in rheumatologists' encounters with their patients who have rheumatoid arthritis. Med Care 1991;29(8):799–812. doi:10.1097/00005650-199108000-00012
11. Homik JE, Suarez-Almazor ME. Clinical guidelines: a must for rheumatology? Best Pract Res Clin Rheumatol 2000;14(4):649–661. doi:10.1053/berh.2000.0105
12. Clinical practice guidelines. American College of Rheumatology. Accessed April 27, 2025. https://rheumatology.org/clinical-practice-guidelines
13. Coutinho AJ, Levin Z, Petterson S, et al. Residency program characteristics and individual physician practice characteristics associated with family physician scope of practice. Acad Med 2019;94(10):1561–1566. doi:10.1097/ACM.0000000000002838

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1277–1286
DOI 10.1002/acr.80002
© 2026 American College of Rheumatology.

American College of Rheumatology
Empowering Rheumatology Professionals

Utilization of Rehabilitation Services Among People With Common Autoimmune Rheumatic Diseases: A Systematic Review

Joel R. Thompson,[1] Keri Geinosky,[2] Josh C. Torrey,[3] Astia Allenzara,[1] and Louise M. Thoma[1]

Objective. Rehabilitation services, including physical and occupational therapy, are frequently recommended in the management of many autoimmune rheumatic diseases (ARDs), yet utilization remains unclear. This systematic review aimed to evaluate how frequently people with common ARDs utilize rehabilitation services.

Methods. We conducted a systematic review of studies published through December 2023 that reported rehabilitation utilization (percentage utilization and/or number of visits) among individuals with ARDs. PubMed and CINAHL were searched. Two reviewers independently screened studies, and data were extracted and summarized by disease, discipline, reporting period, and country.

Results. Of 11,591 records identified, 113 studies met inclusion criteria, and 86 were included in the final analysis. These studies were published between 1991 and 2023, with data representing 22 countries. Rehabilitation utilization was most frequently reported for rheumatoid arthritis (n = 59), axial spondyloarthritis (n = 25), and systemic sclerosis (n = 20). Percentage utilization rates ranged from 0% to 100%, and annual visit counts ranged from 1 to 62.1, varying widely across and within disease types, reporting periods, and countries. Physical therapy was more commonly reported and utilized than occupational therapy. Utilization rates were generally higher in European countries compared to North America.

Conclusion. There was heterogeneity across studies underscoring wide variability in the use of rehabilitation services. These findings highlight inconsistent integration of rehabilitation into rheumatology care across diseases and health systems. Future research should investigate current barriers and facilitators and inform the development of systematic, equitable, and disease-specific strategies to optimize access and delivery of rehabilitation for people with ARDs.

INTRODUCTION

People with autoimmune rheumatic diseases (ARDs) experience a range of physical burdens, such as joint pain, weakness, stiffness, swelling, and fatigue. These symptoms impose substantial challenges to physical function and limit participation in daily life. Even when disease activity is controlled with medication, individuals with ARDs often still experience pain and impaired physical function.[1,2] Pain and limitations in physical function are among the most potent contributors to worse quality of life among adults with ARDs compared to adults without ARDs.[3,4]

Rehabilitation services, including physical therapy (PT) and occupational therapy (OT) services, provide safe and effective strategies for managing pain, weakness, fatigue, and reduced function commonly experienced by individuals with ARD.[5–7] Accordingly, rehabilitation is often indicated to preserve physical function and is recommended in many clinical practice guidelines for management of ARDs, including rheumatoid arthritis (RA), axial spondyloarthritis (axSpA), juvenile idiopathic arthritis (JIA),

Supported in part by the Rheumatology Research Foundation Medical/Graduate Student Preceptorship (to Drs Geinosky and Thoma). The funding sources had no role in study design, collection or interpretation of data, or the decision to submit.

[1]University of North Carolina at Chapel Hill; [2]Kenny, Reynolds, and Associates Physical Therapy, Chapel Hill, North Carolina; [3]Mass General Brigham, Boston, Massachusetts.

Drs Thompson and Geinosky 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.80002).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80002.

Address correspondence via email to Louise M. Thoma, PT, DPT, PhD, at louise_thoma@med.unc.edu.

Submitted for publication August 27, 2025; accepted in revised form December 23, 2025.

SIGNIFICANCE & INNOVATIONS

- This review presents the first comprehensive summary of rehabilitation utilization across common autoimmune rheumatic diseases, countries, and time.
- Despite recommendations for rehabilitation across several treatment guidelines, rehabilitation utilization varies widely within and across diseases.

and psoriatic arthritis (PsA).[8–14] Despite these guidelines, the extent to which rehabilitation is incorporated into routine rheumatology care varies widely across health systems and countries.

Multidisciplinary care involving physical and occupational therapists may be incorporated from initial diagnosis as part of routine care in some contexts (eg, country, health system).[14] Whereas in other contexts, many people never receive rehabilitation services despite years of disease activity and potential damage.[15] This variability may contribute to differences in actual utilization of rehabilitation services, yet utilization patterns across ARDs remain poorly understood. Understanding these patterns is critical for identifying gaps in access and informing strategies to optimize care. The aim of this systematic review was to evaluate how frequently people with common ARDs utilize rehabilitation services. Specifically, we selected RA, axSpA, PsA, JIA, systemic lupus erythematosus (SLE), systemic sclerosis (SSc), and idiopathic inflammatory myopathies (IIM) to represent a breadth of ARDs with varying manifestations and prevalence.

MATERIALS AND METHODS

Study design. This study was a systematic review conducted using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. The study protocol was developed and registered with PROSPERO (CRD42023492966).

Data source and search strategy. A systematic search of PubMed and CINAHL was performed on December 12, 2023, to include all articles published until that date. To evaluate trends in rehabilitation utilization across ARDs, we selected a broad range of ARDs with varying clinical presentations. We combined expected search terms associated with rheumatic disease, utilization, and rehabilitation. Complete search terms can be seen in Supplementary Table S1. Duplicate records were removed. Citation checking was used to check appropriate studies for additional sources, which yielded the inclusion of one additional study outside of these search terms.

Inclusion criteria and study selection. Studies were included if they met the following criteria: (1) participants had RA, axSpA, PsA, JIA, SLE, SSc, or IIM, or data were reported for a combination of these diseases, which we termed "general rheumatology" and (2) studies reported the utilization of PT, OT, and/or allied health services. Allied health services was understood to include PT and/or OT, with or without other professionals such as speech-language pathologists, social workers, or chiropractors. All abstracts and potentially eligible full-text articles were independently reviewed and screened by two reviewers (KG, JRT, and JT). Disagreements were resolved by a third reviewer (LT) through consensus discussion among the reviewers. We included retrospective and prospective observational studies, including administrative and registry data, and clinical trials. Randomized controlled trials (RCTs) were included only when rehabilitation utilization was reported as a baseline or sample description and was not part of the intervention itself. For example, if rehabilitation utilization was reported for an RCT of a biologic drug, it was included, and all reported groups would be pooled for one rehabilitation utilization estimate for the study.

Data extraction and management. One reviewer (KG, JRT, or JT) independently extracted primary outcomes and descriptive information. A second reviewer (KG, JRT, JT, or AA) confirmed the accuracy of the extracted data.

Outcomes. The primary outcomes were rehabilitation utilization, defined in two ways: percentage utilization and number of visits. Percentage utilization was defined as the number of people who utilized the service as a percentage of the total sample in a given reporting period. Number of visits was defined as the quantity of visits to rehabilitation services among individuals who accessed rehabilitation services in a given reporting period. We extracted data for PT and/or OT services independently when available; if estimates were only reported for a combined category of allied health services, then this was extracted. For each utilization estimate, we also extracted the country where the sample was recruited, the year of publication and earliest year of data collection (when reported), and the reporting period for which utilization was evaluated (eg, utilization within a one-year period). If a study only reported utilization estimates for subgroups of the total sample, the estimates were pooled to provide a single estimate per disease per country per reporting period. If a study reported rehabilitation utilization for a shorter reporting period (eg, one year) and ever, then both were extracted. If a study reported utilization over sequential periods (eg, annually over 10 sequential years), only the earliest period was extracted. For studies that reported both the percentage of participants who utilized rehabilitation services and the average number of visits for the entire group (inclusive of nonusers), average visits per user were derived by dividing the reported group mean by the utilization proportion. This approach allowed for standardized reporting of visit counts among individuals who accessed rehabilitation services during the reporting period. Other descriptive characteristics were extracted as available, including study design, data source,

rehabilitation setting, demographics, socioeconomic status, and clinical characteristics.

Assessment of risk of bias. Due to the unique nature of our study investigating estimates of rehabilitation utilization, existing critical appraisal tools were not applicable to the question. As a result, we adapted the JBI critical appraisal tool for prevalence studies[16] and the Newcastle-Ottawa Quality Assessment Form for Cohort Studies[17] to assess the risk of bias. Specific questions and modifications are available in Supplementary Table S2. One reviewer (KG, JRT, or JT) independently assessed for risk of bias. A second reviewer (KG, JRT, or JT) confirmed the accuracy of the assessment. Disagreements were resolved through consensus discussion among all reviewers and the senior author (LT).

Data synthesis. Data were synthesized by discipline, disease, reporting period, and country. Discipline was categorized as PT only (PT), OT only (OT), or a combination of PT, OT, and/or any allied health services (any rehabilitation). Diseases were categorized as listed previously. If a study reported data on samples with different diseases independently, each estimate was used independently. If the study sample combined eligible diseases and/or did not specify the disease, the estimates were synthesized within general rheumatic disease. Reporting period was reduced into four categories (less than six months, one year, two to four years, and ever [ie, five or more years]) to facilitate synthesis and interpretation. If a study reported estimates for multiple countries independently, then an estimate per country was reported to allow for potential synthesis based on location. We calculated descriptive statistics (median, range) of rehabilitation utilization within each disease and reporting period category. This was not weighted by number of participants given the vast heterogeneity between how samples were identified (eg, administrative data, single center cohort, registry), how rehabilitation services were described (eg, self-report, insurance claims, medical record review), and the resulting differences in risk of bias.

RESULTS

The search strategy identified 11,591 references. After removing duplicates, 10,382 records underwent title and abstract screening, 566 were included for full-text review, and 113 studies met the inclusion criteria for data extraction (Figure 1). After excluding 5 studies with duplicate samples, 108 unique studies were included in the analysis. The included studies were published between 1991 and 2023 and represented data from 21 countries. Additional descriptive characteristics of the included studies are described in Supplementary Table S3. Of the 108 studies, 22 studies reported an appropriate outcome (number of rehabilitation visits); however, they reported mean visit counts of the entire sample, including unreported numbers of nonusers, which was unexpected. Given the data are challenging to interpret without the context of number of users, these studies are only reported in Supplementary Table S4. The remaining results synthesize data extracted from the remaining 86 studies that reported data on a primary outcome.

Risk of bias. A summary of the results of the risk of bias assessment is found in Figure 2. Overall, 7 studies (8%) were rated as having low risk of bias, 20 studies (23%) were rated as having moderate risk of bias, and 59 studies (69%) were rated as having high risk of bias. The most frequent source of potential bias was related to self-reported outcome assessment (Figure 2). Additional risk of bias was related to sampling or population demographics, primarily due to convenience sampling of small geographic regions. A full report of the risk of bias assessment can be found in Supplementary Table S5.

Percentage utilization. Table 1 summarizes the number of studies (n) and the median and range for the percentage utilization estimates within each disease by reporting period and discipline. Most estimates reported percentage utilization in a one-year period (n = 68), followed by ever (n = 26), a less than or equal to six-month period (n = 24), and a two- to four-year period (n = 10). Percentage utilization per study is detailed in Supplementary Table S6.

Percentage utilization by disease. Percentage utilization was most frequently reported among samples with RA (n = 59), followed by SSc (n = 25), axSpA (n = 20), SLE (n = 12), PsA (n = 5), IIM, JIA, and general RD (n = 4). Across and within diseases, the variability of rehabilitation utilization percentage estimates was broad. For example, estimates for rehabilitation utilization ranged from 0% to 100% among samples of adults with RA. Similarly, percentage utilization ranged from 0.3% to 100% across the remaining diseases (Table 1). There was also variability between median estimates between diseases: RA had an annual median utilization estimate of 26%, whereas the annual median utilization estimate for axSpA was 37.5%.

Percentage utilization by discipline. Percentage PT utilization was most frequently reported (n = 73), followed by any rehabilitation (n = 30) and OT utilization (n = 25). PT utilization was reported by at least one study for every disease, whereas OT utilization and any rehabilitation utilization were reported for a limited number of diseases. Percentage PT utilization was consistently higher than percentage OT utilization. Among samples reporting both PT and OT utilization (n = 24), percentage PT utilization was higher than percentage OT utilization for 23 of 24 samples; in fact, percentage PT utilization was at least double that of OT in 17 of the samples. This trend was consistent across countries, reporting periods, and diseases. Most OT studies reported percentage utilization <17% (range 0.3%–46%), whereas the median percentage PT utilization was 33%. The highest percentage PT utilization was reported in axSpA, IIM, and SSc, whereas the highest percentage OT utilization was reported in RA, IIM, and SSc.

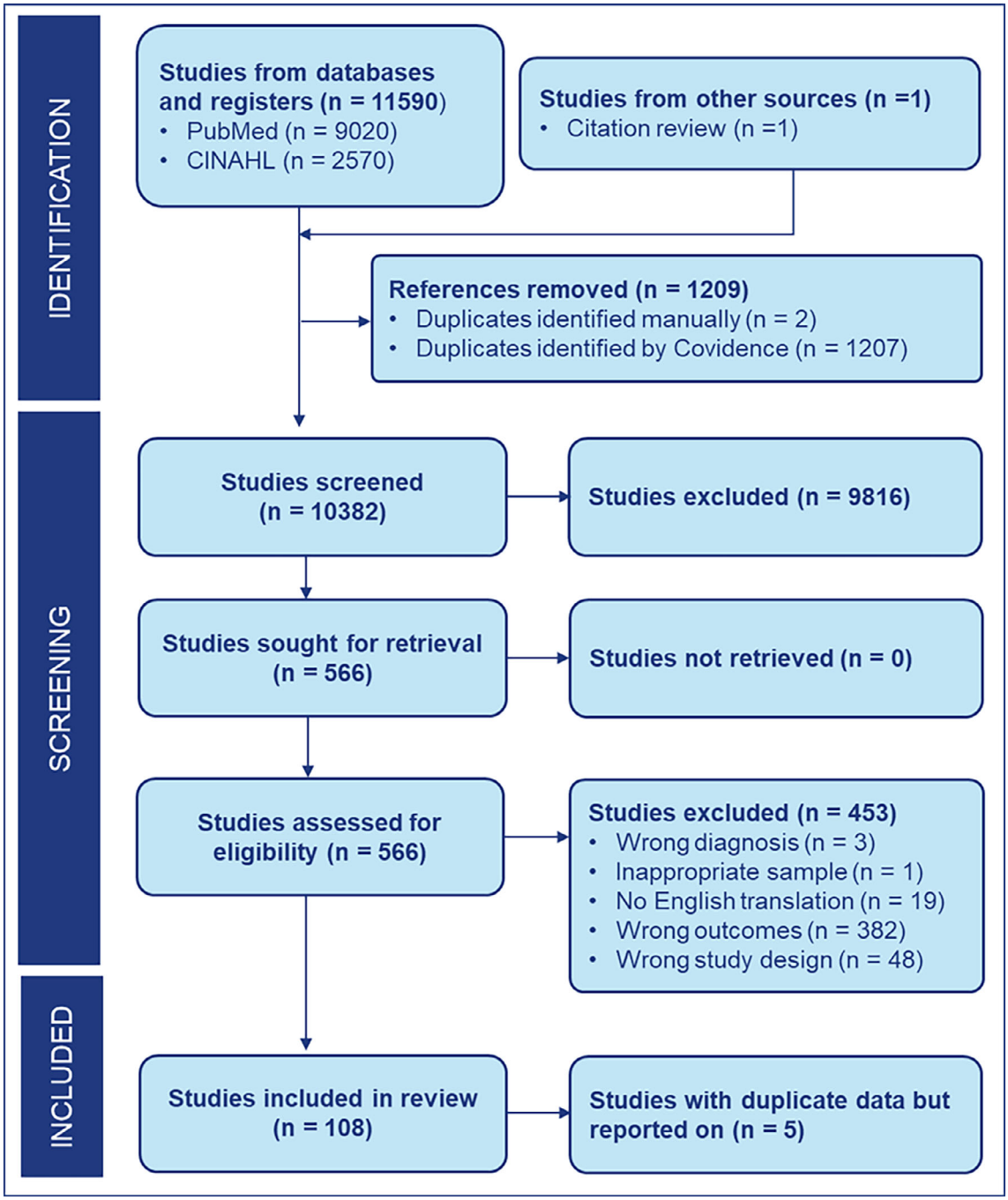

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the selection process of included studies.

Despite any rehabilitation representing a combination of PT, OT, and/or other allied services, median estimates for any rehabilitation were lower than isolated PT estimates across all diseases.

Percentage utilization over time. Studies reporting percentage utilization represented data collection periods beginning as early as 1983 and as recently as 2021. Two studies collected data before 1990, 23 studies collected data starting in the 1990s, 50 studies collected data in the first decade of the 2000s, 49 studies collected data in the 2010s, and 4 studies collected data since 2020. Most studies that collected data in the period before the 2000s were for RA (n = 21), followed by axSpA (n = 3), and a single study each for SSc and SLE. There were no clear trends in percentage utilization over time; however, synthesis was limited by the data heterogeneity.

Percentage utilization across countries. Percentage utilization was reported for samples from 21 countries, with studies from the Netherlands (n = 26) and the United States (n = 23) being the most prevalent. There were more studies published in Europe

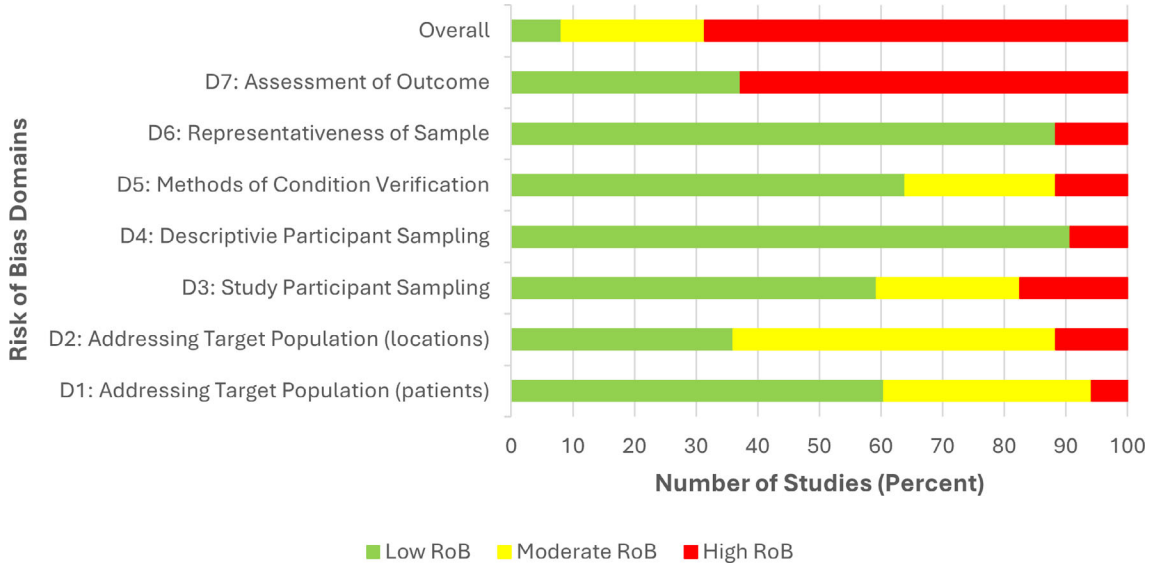

Figure 2. Summary of RoB assessment across included studies. RoB, risk of bias. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.80002/abstract.

(n = 84) than North America (n = 33). A smaller number of studies were conducted in Asia (n = 7) and South America (n = 2), and both South American studies were from Colombia. No studies reported data from African countries. In general, the median percent utilization was higher in European countries compared to North America. For example, the median annual PT utilization for RA was 33.3% (n = 13, range 3.5%–54.1%) in Europe and 12.4% (n = 4, range 9.6%–15.3%) in North America. However, it is important to note great variability in Europe, with a median annual PT utilization for RA of 51.2% (n = 3, range 47.0%–54.1%) in Germany and 25.0% (n = 4, range 3.5%–40.0%) in the Netherlands. Further synthesis was limited by the data heterogeneity.

Number of visits. Table 2 summarizes the number of studies reporting number of visits and the median and range for the number of visits by discipline, disease, and reporting period. Visit counts were only reported over six months or one year. PT visit counts were most frequently reported (n = 23), followed by any rehabilitation (n = 10) and OT visits (n = 7). PT visit counts were available for samples with RA, axSpA, SSc, SLE, and PsA. Similar to percentage utilization, more PT visits were reported than any rehabilitation or OT. Most included studies (18 of 30) reported data from Europe, with studies from the Netherlands (n = 8) and the United Kingdom (n = 5) being most prevalent. Additionally, nine studies reported data from North America, two studies reported data from South America, and one study reported data from Asia. The highest annual visit counts were reported in the Netherlands, followed by Spain and France. In fact, six of the seven highest annual visit counts were reported in the Netherlands, ranging from 38.6 to 62.1 annual visits across axSpA, RA, SSc, and SLE. The highest US annual visit counts were 16.8 annual visits for SLE and 16.7 annual visits for RA. However, high heterogeneity was seen within countries in visit counts, such as annual PT visit counts for people with SSc in the Netherlands varying from 1.3 visits per year to 52 visits per year. Visit counts per study are detailed in Supplementary Table S7.

DISCUSSION

This systematic review provides the first comprehensive synthesis of rehabilitation utilization across multiple ARDs and offers a structured resource organized by discipline, disease, country, and reporting period (Supplementary Table S3). A key strength of this work was the breadth of diseases and contexts captured, enabling broad cross-national observations and exploration of patterns. This study addressed a long-standing gap in the literature and serves as a reference point for future investigations and policy discussions. Overall, rehabilitation utilization was heterogenous and inconsistent across countries and diseases. There was also substantial heterogeneity in study designs, data collection methods, and reporting periods, which precluded meta-analysis and limited the ability to directly compare utilization rates.

Given the wide heterogeneity in utilization, a more nuanced understanding of how care is realized within health systems was necessary to interpret these findings. Levesque et al[18] and Aday and Andersen[19] described health care utilization as realized access that is shaped by the interaction between supply-side factors (such as service availability, affordability, and integration into care pathways) and demand-side factors (such as an individual's capacity, resources, and willingness to engage with care). These frameworks offer a multilevel lens that captures how system-level structures, provider practices, and patient-level capacities collectively determine access and utilization. Within their models, low or

Table 1. Rehabilitation percentage utilization across disciplines, diseases, and time periods*

	<6 mos		1 y		2–4 y		Ever (5+ y)	
	n	Median (range), %	n	Median (range), %	n	Median (range), %	n	Median (range), %
Physical therapy (n = 73)								
RA	5	16.0 (2.0–29.8)	17	26.0 (3.5–54.1)	3	50.0 (7.4–50.0)	5	44.7 (15.5–53.0)
axSpA	2	19.2 (18.0–20.3)	13	37.5 (13.7–80.0)	–	–	2	74.9 (64.7–85.0)
SSc	–	–	6	40.0 (22.0–57.5)	2	33.0 (3.0–63.0)	2	58.1 (34.8–81.3)
SLE	1	9.0	3	21.9 (21.0–26.0)	–	–	1	61.0
PsA	–	–	4	27.6 (8.8–39.7)	–	–	–	–
Myositis	–	–	2	46.3 (29.1–63.4)	–	–	1	79.0
JIA	2	21.6 (19.0–24.1)	–	–	–	–	–	–
General	1	16.0	1	60.0	–	–	–	–
Occupational therapy (n = 25)								
RA	2	15.1 (13.2–17.0)	3	10.0 (3.0–17.0)	2	15.5 (14.0–17.0)	4	18.9 (2.9–30.1)
axSpA	–	–	1	6.0	–	–	1	15.9
SSc	–	–	4	13.6 (8.9–16.0)	1	0.3	2	17.3 (14.1–20.4)
SLE	–	–	1	8.0	–	–	1	22.0
Myositis	–	–	1	46.3	–	–	–	–
General	1	13.0	1	25.0	–	–	–	–
Any rehabilitation (n = 30)								
RA	5	8.3 (0–14.0)	7	23.3 (2.6–100)	–	–	6	27.8 (7.1–41.0)
axSpA	–	–	1	16.0	–	–	–	–
SSc	5	23.0 (17.0–43.0)	2	25.8 (12.0–39.6)	–	–	1	14.2
PsA	–	–	1	14.0	–	–	–	–
JIA	–	–	–	–	2	87.6 (75.2–100)	–	–

* This table synthesizes the median and range of utilization percentage across studies. The data are presented by discipline, disease, and the time period for reporting. If a disease was not listed, there were no studies reporting relevant data. axSpA, axial spondyloarthritis; General, disease reported as combined or undifferentiated rheumatic disease; JIA, juvenile idiopathic arthritis; PsA, psoriatic arthritis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SSc, systemic sclerosis.

inconsistent utilizations rates may signal structural or systemic barriers, such as limited availability of services, inadequate provider referral processes, or affordability constraints, rather than a true absence of need.[18,19] Even when strong clinical guidance exists, as in the case of axSpA,[20,21] access barriers may impede consistent integration of rehabilitation into routine care. Thus, interpreting rehabilitation utilization requires moving beyond clinical appropriateness alone and considering how services are organized, delivered, and experienced within specific health care contexts.

Broader health system structures significantly influence whether people can access and engage with services. Differences in national policy, reimbursement models, and health care structures shape the degree to which rehabilitation is integrated into routine care. For example, two studies in this review highlight the impact of France's single-payer health care system on utilization patterns. Girard et al described how reimbursement policies in France facilitate more frequent PT use among people with RA compared to the United States.[22] Similarly, Becetti et al reported that people with SSc in France were more than four times more likely to receive PT or OT services than people in other European nations and the United States, attributing this difference to structural elements of the national insurance model.[23]

In addition to differences in financing, international variability in professional roles complicates cross-national comparisons. In some health care systems, rehabilitation-related services such as exercise counseling or physical activity promotion may be delivered by professionals other than physical or occupational therapists, including physicians, nurses, or health coaches.[12,24,25] These differences in provider responsibilities and service models introduce further heterogeneity and make it challenging to establish global benchmarks for rehabilitation utilization. Cross-country differences in utilization likely reflect variation in system capacity, care integration, and the availability, affordability, or perceived appropriateness of rehabilitation services.

Clinician-level factors also influence realized access to rehabilitation. Provider practices, knowledge, and communication shape whether patients are offered rehabilitation and can act upon such recommendations.[18] Previous studies in rheumatology indicate that rheumatology clinicians have low confidence in applying nonpharmacological care[26,27] and may underestimate patient interest or ability to engage in rehabiliation.[26–28] Many also describe rehabilitation as a low clinical priority[28] and report time constraints that limit opportunities to address rehabilitation during routine visits.[26,27] Although some clinical guidelines provide recommendations for rehabilitation,[20,21] many clinicians report being unaware of guidelines[27] or uncertain as to how to apply them.[29] Individual provider practices and habits may therefore help explain the wide variation in rehabilitation utilization observed in this review. Communication styles that prioritize pharmacologic management or present rehabilitation in vague, nonspecific terms have been shown to reduce patient engagement with

Table 2. Rehabilitation visit count across disciplines, diseases, and time periods*

	<6 mos		1 y	
	n	Median (range)	n	Median (range)
Physical therapy (n = 23)				
RA	4	10.6 (8.0–15.2)	5	28.8 (16.7–50.2)
axSpA	–	–	8	36.3 (15.6–39.6)
SSc	–	–	2	26.7 (1.3–52.0)
SLE	1	10.4	2	30.2 (16.8–43.7)
PsA	–	–	3	2.6 (1.9–30.8)
Occupational therapy (n = 7)				
RA	4	6.8 (5.8–8.2)	–	–
SSc	–	–	2	2.6 (2.1–3.0)
SLE	–	–	1	2.4
Any rehabilitation (n = 10)				
RA	3	1.2 (0.0–8.0)	3	11.5 (1.0–12.6)
axSpA	–	–	1	8.3
SSc	1	9.8	–	–
PsA	–	–	1	9.3
JIA	–	–	1	2.9

* This table synthesizes median and range across studies that reported the average number of visits per participant among participants who used rehabilitation. The data are presented by discipline, disease, and reporting period. If a disease was not listed, there were no studies reporting relevant data. axSpA, axial spondyloarthritis; JIA, juvenile idiopathic arthritis; PsA, psoriatic arthritis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SSc, systemic sclerosis.

rehabilitation.[30] Collectively, these clinician- and profession-level factors act as a critical pathway through which access is either facilitated or constrained, influencing realized utilization even when services are otherwise available.

Utilization may also be shaped by the extent to which rehabilitation services are embedded within specific care models and the degree of continuity across transitions in care. This pattern is evident in studies reporting on JIA, in which utilization rates were notably high: 100% and 75.2% in two cohorts with reporting periods of two and four years, respectively.[31,32] These findings may reflect the presence of structural supports unique to pediatric care, including coordination with schools, family involvement, and multidisciplinary care models that facilitate rehabilitation access. However, one study found that PT use declined from 24.1% to 19.0% when people with JIA transitioned from pediatric to adult rheumatology care.[33] Transitions between pediatric and adult care in chronic disease often introduce gaps and reduced access to supportive services.[34] These disruptions underscore how care transitions can interfere with the reach and continuity of rehabilitation, limiting realized access despite clinical appropriateness.

Intentional integration of rehabilitation into rheumatology care pathways may be a critical determinant of consistent service use across ARDs. Among the studies reporting the highest rates of rehabilitation utilization, a common feature was structured multidisciplinary care models that included rehabilitation. In the United Kingdom, pediatric patients with RD received coordinated services through a pediatric rheumatology multidisciplinary team,[31] whereas in Sweden, the Early Intervention in Rheumatoid Arthritis (TIRA) project incorporated rehabilitation as a core element of early team-based disease management.[35] Similarly, Horváth et al described how Hungary's publicly funded long-term care system supported consistent rehabilitation access through formal policy and care infrastructure.[36] Verstappen et al also documented high utilization associated with proactive early arthritis management strategies.[37] In each case, rehabilitation was not a reactive or discretionary service but a routine element of clinical care, embedded into the treatment pathway. These examples suggest that when rehabilitation is built into the design of care delivery, access becomes less contingent on provider discretion or patient self-advocacy.

It is important to acknowledge that realized access depends not only on the availability of services but also on the patient's capacity and willingness to engage with care. Individual-level factors shape utilization patterns in ways not directly visible through service-level data alone. Although exploring the impact of individual-level factors was beyond the scope of the current review, some factors are known to impact rehabilitation utilization and access. Barriers can include out-of-pocket costs, limited or no insurance coverage, transportation difficulties, or lack of availability in rural or underserved communities.[15,38,39] Health literacy and limited awareness of the role of rehabilitation in managing ARDs may also impair perceived need, whereas competing demands, time constraints, or prioritization of pharmacological treatment can reduce motivation to seek care.[40,41] Help-seeking among people with ARDs is further hindered by symptom misattribution, delays in seeking care until functional decline becomes pronounced, and limited behavioral capacity or confidence to pursue specialized services such as rehabilitation, even when the value of such services is recognized.[42–45]

Additionally, fear of symptom exacerbation, particularly among individuals who associate physical activity with prior flares, may further discourage participation in rehabilitation and exercise-based interventions.[46,47] Together, these individual-level factors restrict key dimensions of access and contribute to variation in rehabilitation utilization.

We intended to synthesize utilization trends over time given the transformation in medical and pharmaceutical management of rheumatic disease over the last several decades. For example, we suspected that the introduction of biologic therapies around 2002 may affect clinician and patient decisions to prioritize rehabilitation services and thus affect utilization. However, limited data availability before 2002 and overall data heterogeneity hindered the ability to observe discernible differences in rehabilitation utilization over time. Although advanced pharmaceutical approaches have reduced disease activity and inflammation, functional limitation remains common, suggesting that need for rehabilitation in rheumatology care continues.[2]

To advance more scalable and equitable models of rehabilitation use, future research must examine how access unfolds across both the health system and patient experience. Few studies in this review provided insight into the processes by which people come to recognize a need for rehabilitation, are referred by rheumatology clinicians, or successfully engage with care. Because of these gaps, it is unclear how system structures like multidisciplinary care models, national rehabilitation policies, or reimbursement systems work to lower barriers and improve access. Evaluating how such structures function in practice and impact outcomes is key to informing adoption in settings with limited resources or historically low utilization. Further, patient and clinician attitudes and behaviors remain a critical but underexplored influence on access. Future investigations should map the full pathway from perceived need to sustained use, capturing both supply- and demand-side considerations[18] that influence rehabilitation utilization and identifying actionable targets to facilitate appropriate rehabilitation use.

There were limitations to this systematic review. We used two databases in the search strategy, and gray literature sources were not systematically searched; thus, some eligible studies may have been missed. Heterogeneity in study design, sampling approaches, data sources, and reporting periods prevented meta-analysis and complicated comparisons across diseases, countries, and disciplines. Moreover, rehabilitation utilization was defined and measured in diverse ways across studies, contributing to variability in reported estimates. Finally, a large proportion of included studies were rated as having high risk of bias, which may influence the accuracy and generalizability of the utilization estimates.

This review offers the first comprehensive synthesis of rehabilitation utilization across ARDs, highlighting wide variation by disease, country, and model of care. Despite its known benefits, rehabilitation remains inconsistently integrated into rheumatology care and may represent an ongoing unmet need for many people.

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 Thoma 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. Andersson MLE, Forslind K, Hafström I; BARFOT Study Group. Patients with early rheumatoid arthritis in the 2000s have equal disability and pain despite less disease activity compared with the 1990s: data from the BARFOT study over 8 years. J Rheumatol 2017;44(6):723–731. doi:https://doi.org/10.3899/jrheum.161235
2. Carpenter L, Barnett R, Mahendran P, et al. Secular changes in functional disability, pain, fatigue and mental well-being in early rheumatoid arthritis. A longitudinal meta-analysis. Semin Arthritis Rheum 2020;50(2):209–219. doi:https://doi.org/10.1016/j.semarthrit.2019.08.006
3. Björk M, Dragioti E, Alexandersson H, et al. Inflammatory arthritis and the effect of physical activity on quality of life and self-reported function: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2022;74(1):31–43. doi:https://doi.org/10.1002/acr.24805
4. Salaffi F, Carotti M, Gasparini S, et al. The health-related quality of life in rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis: a comparison with a selected sample of healthy people. Health Qual Life Outcomes 2009;7(1):25. doi:https://doi.org/10.1186/1477-7525-7-25
5. de Jong Z, Vliet Vlieland TP. Safety of exercise in patients with rheumatoid arthritis. Curr Opin Rheumatol 2005;17(2):177–182. doi:https://doi.org/10.1097/01.bor.0000151400.33899.88
6. Liem SIE, Vliet Vlieland TPM, Schoones JW, et al. The effect and safety of exercise therapy in patients with systemic sclerosis: a systematic review. Rheumatol Adv Pract 2019;3(2):rkz044. doi:https://doi.org/10.1093/rap/rkz044
7. Van Thillo A, Vulsteke JB, Van Assche D, et al. Physical therapy in adult inflammatory myopathy patients: a systematic review. Clin Rheumatol 2019;38(8):2039–2051. doi:https://doi.org/10.1007/s10067-019-04571-9
8. England BR, Smith BJ, Baker NA, et al. 2022 American College of Rheumatology guideline for exercise, rehabilitation, diet, and additional integrative interventions for rheumatoid arthritis. Arthritis Rheumatol 2023;75(8):1299–1311. doi:https://doi.org/10.1002/art.42507
9. 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. doi:https://doi.org/10.1002/acr.24025
10. Onel KB, Horton DB, Lovell DJ, et al. 2021 American College of Rheumatology guideline for the treatment of juvenile idiopathic arthritis: recommendations for nonpharmacologic therapies, medication monitoring, immunizations, and imaging. Arthritis Care Res (Hoboken) 2022;74(4):505–520. doi:https://doi.org/10.1002/acr.24839

11. Singh JA, Guyatt G, Ogdie A, et al. 2018 American College of Rheumatology/National Psoriasis Foundation guideline for the treatment of psoriatic arthritis. Arthritis Rheumatol 2019;71(1):5–32. doi: https://doi.org/10.1002/art.40726
12. Rausch Osthoff AK, Niedermann K, Braun J, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018;77(9):1251–1260. doi: https://doi.org/10.1136/annrheumdis-2018-213585
13. Oldroyd AGS, Lilleker JB, Amin T, et al; British Society for Rheumatology Standards, Audit and Guidelines Working Group. British Society for Rheumatology guideline on management of paediatric, adolescent and adult patients with idiopathic inflammatory myopathy. Rheumatology (Oxford) 2022;61(5):1760–1768. doi:https://doi.org/10.1093/rheumatology/keac115
14. Peter WF, Swart NM, Meerhoff GA, et al. Clinical practice guideline for physical therapist management of people with rheumatoid arthritis. Phys Ther 2021;101(8):pzab127. doi:https://doi.org/10.1093/ptj/pzab127
15. Lane CY, Lo D, Thoma LM, et al. Sociocultural and economic disparities in physical therapy utilization among insured older adults with rheumatoid arthritis. J Rheumatol 2023;50(11):1414–1421. doi: https://doi.org/10.3899/jrheum.2023-0103
16. Munn Z, Moola S, Lisy K, et al. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. Int J Evid-Based Healthc 2015;13(3):147–153. doi:https://doi.org/10.1097/XEB.0000000000000054
17. Wells GSB, O'Connell D, Peterson J, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Updated 2013. https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
18. Levesque JF, Harris MF, Russell G. Patient-centred access to health care: conceptualising access at the interface of health systems and populations. Int J Equity Health 2013;12(1):18. doi:https://doi.org/10.1186/1475-9276-12-18
19. Aday LA, Andersen R. A framework for the study of access to medical care. Health Serv Res 1974;9(3):208–220.
20. Ward MM, Deodhar A, Gensler LS, et al. 2019 update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network recommendations for the treatment of ankylosing spondylitis and nonradiographic axial spondyloarthritis. Arthritis Rheumatol 2019;71(10):1599–1613. doi:https://doi.org/10.1002/art.41042
21. Ramiro S, Nikiphorou E, Sepriano A, et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. Ann Rheum Dis 2023;82(1):19–34. doi:https://doi.org/10.1136/ard-2022-223296
22. Girard F, Guillemin F, Novella JL, et al. Health-care use by rheumatoid arthritis patients compared with non-arthritic subjects. Rheumatology (Oxford) 2002;41(2):167–175. doi:https://doi.org/10.1093/rheumatology/41.2.167
23. Becetti K, Kwakkenbos L, Carrier ME, et al; SPIN Investigators. Physical or occupational therapy use in systemic sclerosis: a scleroderma patient-centered intervention network cohort study. J Rheumatol 2019;46(12):1605–1613. doi:https://doi.org/10.3899/jrheum.181130
24. Stamm T, Hill J. Extended roles of non-physician health professionals and innovative models of care within Europe: results from a web-based survey. Musculoskeletal Care 2011;9(2):93–101. doi:https://doi.org/10.1002/msc.201
25. Bech B, Primdahl J, van Tubergen A, et al. 2018 update of the EULAR recommendations for the role of the nurse in the management of chronic inflammatory arthritis. Ann Rheum Dis 2020;79(1):61–68. doi:https://doi.org/10.1136/annrheumdis-2019-215458
26. Derakhshan MH, Pathak H, Cook D, et al; NASS and BRITSpA investigators. Services for spondyloarthritis: a survey of patients and rheumatologists. Rheumatology (Oxford) 2018;57(6):987–996. doi: https://doi.org/10.1093/rheumatology/kex518
27. Kronbi F, Rkain H, Ez-Zaoui S, et al. Attitudes, practices and perceived barriers toward implementing non-pharmacological management for rheumatoid arthritis among rheumatologists: an online cross-sectional survey. Reumatologia 2024;62(4):250–258. doi: https://doi.org/10.5114/reum/191792
28. Iversen MD, Scanlon L, Frits M, et al. Perceptions of physical activity engagement among adults with rheumatoid arthritis and rheumatologists. Int J Clin Rheumatol 2015;10(2):67–77. doi:https://doi.org/10.2217/ijr.15.3
29. Zimba O, Kocyigit BF, Kadam E, et al. Knowledge, perceptions, and practices of axial spondyloarthritis diagnosis and management among healthcare professionals: an online cross-sectional survey. Rheumatol Int 2024;44(8):1501–1508. doi:https://doi.org/10.1007/s00296-024-05638-w
30. Georgopoulou S, Prothero L, D'Cruz DP. Physician-patient communication in rheumatology: a systematic review. Rheumatol Int 2018;38(5):763–775. doi:https://doi.org/10.1007/s00296-018-4016-2
31. Foster HE, Eltringham MS, Kay LJ, et al. Delay in access to appropriate care for children presenting with musculoskeletal symptoms and ultimately diagnosed with juvenile idiopathic arthritis. Arthritis Rheum 2007;57(6):921–927. doi:https://doi.org/10.1002/art.22882
32. Brunner HI, Taylor J, Britto MT, et al. Differences in disease outcomes between Medicaid and privately insured children: possible health disparities in juvenile rheumatoid arthritis. Arthritis Rheum 2006;55(3):378–384. doi:https://doi.org/10.1002/art.21991
33. Mannion ML, Xie F, Baddley J, et al. Analysis of health care claims during the peri-transfer stage of transition from pediatric to adult care among juvenile idiopathic arthritis patients. Pediatr Rheumatol Online J 2016;14(1):49. doi:https://doi.org/10.1186/s12969-016-0107-3
34. Healy EW, Piracha NZ. Evaluating the transition of adolescents and young adults with palliative care needs from pediatric to adult care. Health Care Transit 2024;2:100072. doi:https://doi.org/10.1016/j.hctj.2024.100072
35. Hallert E, Husberg M, Jonsson D, et al. Rheumatoid arthritis is already expensive during the first year of the disease (the Swedish TIRA project). Rheumatology (Oxford) 2004;43(11):1374–1382. doi:https://doi.org/10.1093/rheumatology/keh324
36. Horváth CZ, Sebestyén A, Österle A, et al. Economic burden of long-term care of rheumatoid arthritis patients in Hungary. Eur J Health Econ 2014;15(suppl 1):131–135. doi:https://doi.org/10.1007/s10198-014-0601-9
37. Verstappen SM, Jacobs JW, van der Heijde DM, et al. Utility and direct costs: ankylosing spondylitis compared with rheumatoid arthritis. Ann Rheum Dis 2007;66(6):727–731. doi:https://doi.org/10.1136/ard.2006.061283
38. Sharpe JA, Thackeray A, Fritz JM, et al. Patients' use of physical therapy for lower back pain: A qualitative study. Musculoskelet Sci Pract 2021;56:102468. doi:https://doi.org/10.1016/j.msksp.2021.102468
39. Felter CE, Zalewski K, Jermann R, et al. Rural health: the dirt road less traveled. Phys Ther 2022;102(11):pzac112. doi:https://doi.org/10.1093/ptj/pzac112
40. Sharpe JA, Shapiro A, Goode A, et al. Patient choice after referral to physical therapy for lower back pain and knee osteoarthritis. Arch Rehabil Res Clin Transl 2023;5(1):100243. doi:https://doi.org/10.1016/j.arrct.2022.100243
41. Liem SIE, van Leeuwen NM, Vliet Vlieland TPM, et al. Physical therapy in systemic sclerosis: the patient perspective. Arthritis Care Res (Hoboken) 2023;75(1):145–151. doi:https://doi.org/10.1002/acr.24741

42. Feldman DE, Bernatsky S, Lévesque JF, et al. Access and perceived need for physical and occupational therapy in chronic arthritis. Disabil Rehabil 2010;32(22):1827–1832. doi:https://doi.org/10.3109/09638281003734417
43. Koehn S, Jones CA, Pham A, et al. Identifying the need for care in rheumatoid arthritis: a Candidacy 2.0 analysis of lived experiences. Soc Sci Med 2025;374:118040. doi:https://doi.org/10.1016/j.socscimed.2025.118040
44. Xiang L, Yoon S, Low AHL, et al. Social cognitive theory to improve symptom appraisal and help-seeking among patients with autoimmune rheumatic diseases: a qualitative study. Int J Rheum Dis 2023;26(11):2258–2266. doi:https://doi.org/10.1111/1756-185X.14923
45. Xiang L, Yoon S, Low AHL, et al. Symptom appraisal and help-seeking among patients with autoimmune rheumatic diseases: a qualitative study. J Rheumatol 2023;50(12):1624–1632. doi:https://doi.org/10.3899/jrheum.2023-0086
46. Baxter S, Smith C, Treharne G, et al. What are the perceived barriers, facilitators and attitudes to exercise for women with rheumatoid arthritis? A qualitative study. Disabil Rehabil 2016;38(8):773–780. doi:https://doi.org/10.3109/09638288.2015.1061602
47. Karaca NB, Arin-Bal G, Sezer S, et al. Physical activity, kinesiophobia, pain catastrophizing, body awareness, depression and disease activity in patients with ankylosing spondylitis and rheumatoid arthritis: a cross-sectional explorative study. Musculoskeletal Care 2024;22(4):e1953. doi:https://doi.org/10.1002/msc.1953

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1287–1297
DOI 10.1002/acr.80024
© 2026 American College of Rheumatology.

American College of Rheumatology
Empowering Rheumatology Professionals

Major US Formulary Change Triggers, but Does Not Sustain, Increases in Biosimilar Adalimumab Prescribing

Eric T. Roberts, PhD, MPH,[1] Gabriela Schmajuk, MD, MAS,[1,2] Mackenzie Clark, PharmD,[1] Diana Ung, PharmD,[1] and Jinoos Yazdany, MD, MPH[1,3]

Objective. CVS Caremark changed their national commercial formularies in 2024 to prefer adalimumab-adaz over originator. We assessed the formulary change's effect on biosimilar adalimumab prescribing in US community rheumatology practices.

Methods. Data come from Rheumatology Informatics System for Effectiveness, a registry of electronic health record data from participating community rheumatology practices nationwide. We used interrupted time series to analyze the effect of the CVS formulary change on the proportion of patients prescribed adalimumab-adaz or other adalimumab biosimilars each month between January 2023 and September 2024 stratified by payor and among new users.

Results. We included 22,079 patients. Among those commercially insured, prescriptions for adalimumab-adaz went from nearly 0% to more than 25% following the CVS formulary change. In the next month, prescriptions for adalimumab-adaz dropped 12.3%. Prescriptions for other biosimilars increased slowly and linearly over the study period. Among patients with Medicare, we observed a similarly timed, but much smaller, increase in prescribing for both adalimumab-adaz and other biosimilar molecules without a large subsequent decrease. Among new users, prescriptions for adalimumab-adaz increased from 0% to 24% but was followed by a nearly 14% decrease in prescriptions in the first month after the formulary change. Other biosimilar prescriptions increased 0.6% per month throughout the period.

Conclusions. Although biosimilar adalimumab prescriptions rose sharply after a major national formulary change, the increase was not sustained. These findings highlight that persistent barriers, including the nocebo effect and the administrative burden of switching, may continue to limit biosimilar adoption and contribute to the loss of substantial cost savings in the US market.

INTRODUCTION

Biologic drugs have revolutionized patient care, but their high costs strain patients' and health care's finances. In 2021, biologics accounted for 46% of spending on medication in the United States, totaling $260 billion. Even with the introduction of biosimilars, spending on biologics has increased by 12.5% annually since 2017, far outpacing growth on small molecule drugs (1.3%).[1] Among biologics, adalimumab stands out as the highest-grossing drug in history. Between originator adalimumab's market launch and introduction of biosimilars (delayed by approximately six years), the list price of adalimumab increased 470%,[2] and delayed market entry is estimated to have cost Medicare more than $2 billion.[3]

The reasons for the sluggish US biosimilars market have been extensively discussed, with many commentators pointing to the legal obstacles that prevent or delay market entry.[4–7] In 2023, patent exclusivity for adalimumab expired, and nine biosimilars were introduced to the US market (with a 10th launched in May 2024). Uptake in the first year of availability, however, was very low (~3%),[8,9] underscoring the complex network of barriers to successful uptake in the United States beyond market entry. These include formulary coverage,[10,11] unequal reimbursements,[12,13] a reticence

Supported by the Agency for Healthcare Research and Quality (grant R01-HS-028024) and National Institute for Arthritis and Musculoskeletal and Skin Diseases, NIH (grant K24-AR-074534). The data presented were supported by the American College of Rheumatology's Rheumatology Informatics System for Effectiveness registry. The views expressed herein represent those of the authors and do not necessarily represent the views of the American College of Rheumatology.

[1]University of California, San Francisco; [2]San Francisco Veterans Affairs Medical Center, San Francisco, California; [3]Zuckerberg San Francisco General Hospital, San Francisco, California.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80024.

Address correspondence via email to Eric T. Roberts, PhD, MPH, at Eric.roberts@ucsf.edu.

Submitted for publication August 19, 2025; accepted in revised form February 9, 2026.

SIGNIFICANCE AND INNOVATIONS

- The high costs of biologics is a significant strain on patients and the US health care system; savings and expanded access from biosimilars have yet to reach levels seen in other nations. This paper evaluates biosimilar prescribing dynamics in response to a national payer-driven formulary change to assess best practices for future biosimilar rollouts.
- Overall, 19% of patients initiating adalimumab-adaz switched therapies, with patients preferentially switching back to the drug they started with.
- Although biosimilar adalimumab prescribing rose dramatically in early 2024, corresponding to the formulary change, some of these gains were short-lived; more than 40% of patients were not taking the biosimilar within months.
- The nocebo effect and the administrative burden associated with changing drugs will prevent and delay biosimilar uptake until we institute policies to address these problems.

to switch patients despite equal safety and efficacy,[14–16] the nocebo effect (a phenomenon whereby negative expectations regarding a treatment lead to the perception or experience of worsened symptoms) among patients forced to switch,[17] interchangeability,[18] and removal or changes to patient copay or support programs.[19,20] Despite the promise of cost savings and expanded access, the real-world impact of biosimilar introduction, particularly for commonly used drugs like adalimumab, remains poorly understood.

In the United States, pharmacy benefit managers (PBMs) are intermediaries who manage contractual relationships among employers, insurers, pharmacies, and pharmaceutical manufacturers to, among other tasks, administer prescription drug formularies. Formularies determine which drugs are covered on which tier and specify which use management tools are in place.[21,22] PBMs administer drug benefits for more than 275 million Americans, and the biggest three, CVS Caremark, Express Scripts, and Optum Rx, processed nearly 80% of prescription claims in the United States in 2023 and 2024.[21,23] CVS Caremark was the largest PBM in the United States in 2023 and second largest in 2024, processing 34% and 27% of claims, respectively.[23] In January 2024, CVS Caremark announced the removal of adalimumab originator from its national commercial formularies in favor of adalimumab-adaz effective April 2024.[24–26] In this study, we assess whether this decision changed prescribing of adalimumab-adaz and other biosimilar adalimumab molecules in community rheumatology practices nationwide.

METHODS

Data source. Data come from the American College of Rheumatology's Rheumatology Informatics System for Effectiveness (RISE) registry, which includes electronic health record (EHR) data from approximately 1,100 rheumatologists across 225 practices.[27] Data collection procedures in the RISE are approved by the University of California, San Francisco, and the Western Institutional Review Board.

Study population. We included individuals 18 years or older with Medicare, Medicaid, or private insurance who had at least one visit with a rheumatologist in the RISE and adalimumab bio-originator or biosimilar use, regardless of diagnosis. All bio-originator or biosimilar adalimumab prescriptions recorded in electronic prescription (e-prescription) tables (of any dose) between January 1, 2023 (time of market launch of the first adalimumab biosimilar), and September 30, 2024, were included.

For patients with orders for different drugs on the same day, we assigned medication use based on their subsequent record (ie, if their next order was a biosimilar, then we considered the ambiguous order a biosimilar). This approach is based on our clinical experience; providers may prescribe both biosimilar and bio-originator on the same day to see what insurance companies will approve; the subsequent prescription is therefore the best indication of what was received by the patient.

Cohort construction. To report the demographic and clinical characteristics of the sample, we defined an "ever use" cohort comprised of the 22,079 individuals who met the inclusion criteria above. The "new user" cohort was restricted to the 12,945 individuals with a first administration of adalimumab or a biosimilar in the RISE on or after January 1, 2023, and at least one visit in the 365 days before their first administration of adalimumab (to ensure observability) with no adalimumab use.

Covariates. The following covariates were extracted from the RISE: patient insurance coverage, sex, and age. Race and ethnicity were self-reported and described here based on Medicare categories. We tested for differences in uptake of biosimilars by race and ethnicity to assess for health disparities in prescribing practices. We included socioeconomic status using the Area Deprivation Index (ADI), a measure of socioeconomic disadvantage at the patient residential Census Block Group level, which is calculated using 17 indicators from the American Community Survey that encompass income, education, employment, and housing conditions.[28] The patient's primary diagnosis was defined by two International Classification of Diseases (ICD) codes at least 30 days apart between January 2022 and September 2024 for psoriatic arthritis (PSA), ankylosing spondylitis (AS), rheumatoid arthritis (RA), or other. A clinically adjudicated hierarchy was applied to patients with ICD codes for more than one disease so that patients were only assigned to one diagnosis category: PSA > AS > RA > other.[29]

Statistical analyses. *Descriptives.* We tabulated the number of patients with a prescription for each drug in 2023 and in 2024. We tested for differences in ever use of a biosimilar between the years as well as ever use in 2023 or 2024 combined across insurance coverage, demographics, and diagnosis using standardized mean differences (SMDs). The SMD gives insight into the magnitude of differences among groups rather than a probability of statistical significance, which is particularly useful with large sample sizes. The stddiff package in Stata was used, which uses a multivariate Mahalanobis distance method to generalize the standardized difference metric to handle a multinomial sample.[30] SMDs <0.2 were considered as "very small," 0.2 to 0.5 "small," 0.5 to 0.8 "medium," and >0.8 "large."[31] Missing data were included as an unknown category.

Interrupted time series by payor and new users. We used an interrupted time series (ITS) to model the effect of the CVS formulary change.[32] ITS offers a robust quasi-experimental research design that can estimate the causal effect of an intervention on a population when randomization is not feasible or practical and data are available at multiple time points pre- and postintervention. It has been used to evaluate market introductions of biosimilars,[12] the effect of national policy changes on biosimilar uptake,[33] and the effect of a national EHR-based dashboard to support safe prescribing of biologics.[34]

For each insurance payor (private; Medicare) and new users, we modeled the proportion of prescriptions for each month during the study period for (1) adalimumab-adaz, and (2) all other biosimilars combined unadjusted. Initial models with the interruption in April 2024 did not fit the observed data well. Therefore, following recommendations, we revised our modeling strategy using visual inspection of the observed data to help define the trends.[35] We observed the following three trends and note which prescriptions they applied to parenthetically: (1) an anticipatory response to the new policy beginning in February 2024 (private: adalimumab-adaz; Medicare: adalimumab-adaz and other biosimilars; new users: adalimumab-adaz), (2) a lagged decrease in prescribing occurring in May 2024 (private: adalimumab-adaz; new users: adalimumab-adaz), and (3) no response to the policy change at any time point (private: other biosimilars; new users: other biosimilars). Accordingly, our final models include interruptions at February 2024 and May 2024 as appropriate, and in cases of no response, we present linear models to keep results parsimonious. To account for autocorrelation, we estimated Newey-West standard errors in ITS models with the Stata itsa package, and we used Yule-Walker estimation for linear regressions in SAS proc autoreg.

Pathways of adalimumab-adaz use. ITS graphs revealed a large drop-off in use of adalimumab-adaz after an initial three months of increasing use. To characterize this, we mapped pathways of use for all patients who ever received adalimumab-adaz, including what drug they initiated and their status at the end of the study. We categorized patients as starting as new users (ie, their first adalimumab molecule was adalimumab-adaz based on our new user criteria outlined above) or if they switched from originator or biosimilar adalimumab (determined by the prescription closest in time to the index administration of adalimumab-adaz). A total of 27 patients with their first visit in the RISE corresponding to their index prescription were omitted because we could not reliably determine if they were a new user or switching. Patient's end status was determined as follows: if a patient had a prescription for a different adalimumab molecule or another biologic disease-modifying antirheumatic drug (bDMARD) or targeted synthetic disease-modifying antirheumatic drug (tsDMARD) within six months of their last adalimumab-adaz prescription, we classified them as switchers. If a patient's last prescription for adalimumab-adaz was within six months of September 2024 and they did not have a prescription for another bDMARD/tsDMARD, they were considered to still be using adalimumab-adaz. If a patient's last prescription was more than six months from September 2024 and they had no other prescription, they were considered to have stopped biologic therapy at least temporarily. It is possible, given the limited follow-up time, that these stops were extended interruptions to therapy induced by administrative barriers or uncertainty over costs. To ensure observability, patients were required to have at least one visit with a RISE rheumatologist after their last adalimumab-adaz prescription. This requirement excluded five patients, suggesting the potential bias due to moving or changing rheumatologists is minimal. Results are presented as a Sankey diagram. Data were managed and cleaned in SAS version 9.4; linear regressions were completed with proc autoreg; additional analyses were completed using the stddiff and itsa packages in Stata/MP version 18.

RESULTS

Table 1 displays the demographic and clinical characteristics of the 22,079 patients included in our ever use cohort. Our sample was 68% female, 65% non-Hispanic White, 38% aged 51 to 64 years, and 28% aged 65 years or older. RA was the most common diagnosis (54%), and the majority of patients had private insurance (67%), followed by Medicare (27%) and Medicaid (6%). Overall, 3,486 patients (15.8%) were prescribed an adalimumab biosimilar at any point during the study period. Variables with SMDs indicating appreciable imbalance between biosimilar users versus originator users included calendar year, insurance type, and age group, although age and insurance were collinear. There were only minor differences in sex, race and ethnicity, ADI, and diagnosis. Biosimilar prescriptions were more common among privately insured individuals (16.4%) than those with Medicaid (8.2%) or Medicare (6.5%). In 2023, fewer than 4% of patients received a biosimilar, increasing to nearly 25% in the first nine months of 2024 (Table 2). Among biosimilar prescriptions in 2024, the most

Table 1. Demographic and clinical characteristics of patients prescribed adalimumab or its biosimilars in the RISE 2023 to 2024*

Characteristic	Total (n = 22,079), n (%)	No biosimilar (n = 18,593), n (%)	Ever biosimilar (n = 3,486), n (%)	SMD[a]
Insurance				0.43
Medicaid	1,392 (6.3)	1,259 (90.45)	133 (9.55)	
Medicare	6,012 (27.2)	5,537 (92.10)	475 (7.90)	
Private	14,675 (66.5)	11,797 (80.39)	2,878 (19.61)	
Year[b]				1.08
2023	15,643 (55.5)	15,070 (96.34)	573 (3.66)	
2024 (first nine months)	12,534 (44.5)	9,406 (75.04)	3,128 (24.96)	
Age				0.35
18–30 years	1,194 (5.4)	1,000 (83.75)	194 (16.25)	
31–50 years	6,174 (28)	4,997 (80.94)	1,177 (19.06)	
51–64 years	8,487 (38.4)	6,927 (81.62)	1,560 (18.38)	
>64 years	6,224 (28.2)	5,669 (91.08)	555 (8.92)	
Sex				0.11
Female	15,026 (68.1)	12,812 (85.27)	2,214 (14.73)	
Male	7,053 (31.9)	5,781 (81.97)	1,272 (18.03)	
Race and ethnicity				0.08
Black	1,520 (6.9)	1,284 (84.47)	236 (15.53)	
Asian	345 (1.6)	271 (78.55)	74 (21.45)	
Hispanic	830 (3.8)	701 (84.46)	129 (15.54)	
Other/mixed	169 (0.8)	145 (85.80)	24 (14.20)	
Unknown	4,792 (21.7)	4,113 (85.83)	679 (14.17)	
White	14,423 (65.3)	12,079 (83.75)	2,344 (16.25)	
ADI percentiles				0.16
25th (high SES)	4,519 (20.5)	3,628 (80.28)	891 (19.72)	
50th	6,364 (28.8)	5,344 (83.97)	1,020 (16.03)	
75th	5,723 (25.9)	4,898 (85.58)	825 (14.42)	
99th (low SES)	4,063 (18.4)	3,490 (85.90)	573 (14.10)	
Missing	1,410 (6.4)	1,233 (87.45)	177 (12.55)	
Diagnosis				0.05
RA	11,926 (54.0)	10,087 (54.30)	1,839 (52.80)	
PSA	2,261 (10.2)	1,902 (10.20)	359 (10.30)	
AS	2,881 (13.0)	2,380 (12.80)	501 (14.40)	
Other	5,011 (22.7)	4,224 (22.70)	787 (22.60)	

* ADI, Area Deprivation Index; AS, ankylosing spondylitis; PSA, psoriatic arthritis; RA, rheumatoid arthritis; RISE, Rheumatology Informatics System for Effectiveness; SES, socioeconomic status; SMD, standardized mean difference.
[a] SMDs were calculated using the stddiff package in Stata, which uses a multivariate Mahalanobis distance method to generalize the standardized difference metric to handle a multinomial sample.
[b] Year calculation is based on receipt of biosimilar by patient in that year; all other calculations are ever received in 2023 to 2024 combined.

frequently prescribed drugs were adalimumab-adaz (55.5%) and adalimumab-bwwd (19%; Table 2).

Table 3 presents the model estimates with confidence intervals for our final models, which are depicted graphically in Figures 1A and B and 2. To qualitatively summarize the major trends we observed, for privately insured patients and new users, there were basically no adalimumab-adaz prescriptions until there was a large increase immediately following the announcement of the formulary change followed by a large decrease three months later. Prescriptions for other adalimumab biosimilars in these groups grew steadily over time. Medicare beneficiaries saw modest increases in prescribing for adalimumab-adaz and other biosimilar adalimumab molecules over time before a less pronounced jump immediately following the announcement of the formulary change followed by no large decrease in prescribing.

Figure 1A graphs the proportion of prescriptions among privately insured patients monthly during our study period. There were nearly no prescriptions for adalimumab-adaz from January 2023 to January 2024; prescriptions increased substantially to 17 percentage points (pp) in February 2024. Uptake continued to rise at the rate of more than 4 pp per month through April 2024. However, this increase was short-lived, with biosimilar prescriptions declining 12.3 pp from April to May. Biosimilar use continued to decline thereafter at a rate of 0.9 pp/mo. Prescriptions for the other, less commonly used biosimilars increased steadily at a rate of 0.8 pp/mo over the study period.

Figure 1B graphs the proportion of prescriptions among patients with Medicare monthly during our study period. The pattern of prescriptions over time for adalimumab-adaz and other biosimilars was very similar. There was a small increase in the proportion of prescriptions for a biosimilar each month between January 2023 and January 2024 (0.1 pp for adalimumab-adaz; 0.3 pp for other

Table 2. Count of distinct patients prescribed adalimumab or its biosimilars in the RISE by year*

Version	2023	2024[a]
adalimumab	15,293	10,285
adaz	43	1,824
bwwd	138	624
atto	277	213
adbm	33	172
aqvh	48	131
ryvk	0	117
aacf	20	105
aaty	29	66
fkjp	8	33
afzb	0	2

* RISE, Rheumatology Informatics System for Effectiveness.
[a] Data are available through September 2024.

biosimilars). In February 2024, adalimumab-adaz prescriptions increased 6.9 pp, and other biosimilars increased 2.9 pp; following this peak, both groups declined at a consistent rate of 0.3 pp/mo.

Figure 2 graphs the proportion of prescriptions among new users monthly during our study period. There was a 0.1-pp increase in the proportion of prescriptions for adalimumab-adaz each month between January 2023 and January 2024 but a jump of >19 pp in February 2024. Uptake continued to rise at the rate of 2.4 pp/mo through May 2024. Similar to the pattern observed in the overall cohort, there was a 13.8-pp drop from April to May, with use continuing to decline thereafter at a rate of 0.7 pp/mo through September. Prescriptions for the other biosimilars increased steadily at a rate of 0.6 pp/mo over the study period.

Figure 3 is a Sankey diagram plotting the pathways in and out of use of adalimumab-adaz. Of the total 1,828 patients who were prescribed adalimumab-adaz during the study period, 989 (54.1%) were new users, 775 (42.3%) switched from originator adalimumab, and 64 (3.5%) switched from a different biosimilar adalimumab. By September 2024, 1,045 patients (57.1%) were still using adalimumab-adaz, 440 (24.1%) had discontinued biologic therapy, 168 (9.2%) switched to a different adalimumab molecule, and 175 (9.6%) switched to a different biologic. Overall, patients who switched to adalimumab-adaz were more likely to switch to a different drug than new users (23% vs 15%), with patients preferentially switching back to the drug they started with. Switching from adalimumab-adaz to originator happened more frequently in patients who started with taking originator (9.2%; 71 of 775) compared to new users (5%; 49 of 989), and patients started on biosimilar (3%; 2 of 64). Similarly, patients who started with taking biosimilar adalimumab were more likely to end taking biosimilar adalimumab (6.3%; 4 of 64) than new users (2.1%; 21 of 989) or patients who started with taking originator (2.7%; 21 of 775).

DISCUSSION

In this national study of more than 22,000 US patients treated in community rheumatology practices, we examined the impact of a major formulary change that shifted coverage from originator adalimumab to a biosimilar, adalimumab-adaz. Adalimumab is one of the most prescribed and historically highest-grossing biologics in the United States, making this an important opportunity to evaluate real-world biosimilar adoption. We found that although biosimilar adalimumab prescribing rose dramatically in early 2024, corresponding to the formulary change, some of these gains were short-lived; more than 40% of patients stopped biologic therapy, at least temporarily, within months. Among patients with private insurance, prescriptions for adalimumab-adaz peaked in the month of the formulary change. Similar though less pronounced trends were observed among

Table 3. Estimates from final models of the proportion of prescriptions each month for adalimumab-adaz and other adalimumab biosimilars for patients with private insurance and Medicare and among new users*

Model and parameter	Coefficient	UCI	LCI
Private insurance			
Adalimumab-adaz			
Intercept	−0.19	−0.41	0.03
Time	0.07	0.04	0.10
Level change 1	16.88	13.65	20.11
Trend change 1	4.11	1.48	6.74
Level change 2	−12.34	−18.09	−6.58
Trend change 2	−5.12	−7.72	−2.53
Other biosimilars			
Intercept	−1.98	−2.79	−1.18
Time	0.84	0.77	0.90
Medicare			
Adalimumab-adaz			
Intercept	−0.21	−0.46	0.03
Time	0.08	0.04	0.12
Level change 1	6.87	5.96	7.78
Trend change 1	−0.41	−0.57	−0.24
Other biosimilars			
Intercept	−0.06	−0.44	0.33
Time	0.27	0.19	0.34
Level change 1	2.88	1.59	4.16
Trend change 1	−0.55	−0.76	−0.34
New users			
Adalimumab-adaz			
Intercept	−0.25	−0.55	0.05
Time	0.10	0.05	0.14
Level change 1	18.33	16.02	20.64
Trend change 1	2.30	0.36	4.24
Level change 2	−13.83	−19.15	−8.51
Trend change 2	−3.14	−5.10	−1.19
Other biosimilars			
Intercept	−0.04	−1.41	1.34
Time	0.62	0.51	0.73

* Intercept is the mean proportion of prescriptions for the stated group of biosimilars in the zeroth month of observation (December 2023). Time is the preintervention slope of the change in proportion of prescriptions each month. Level change 1 is the level change in proportion of prescriptions between January 2024 and February 2024. Trend change 1 is the difference in slope between before and after February 2024. Level change 2 is the level change in proportion of prescriptions between April 2024 and May 2024. Trend change 2 is the additional difference in slope after May 2024. LCI, lower confidence interval; UCI, upper confidence interval.

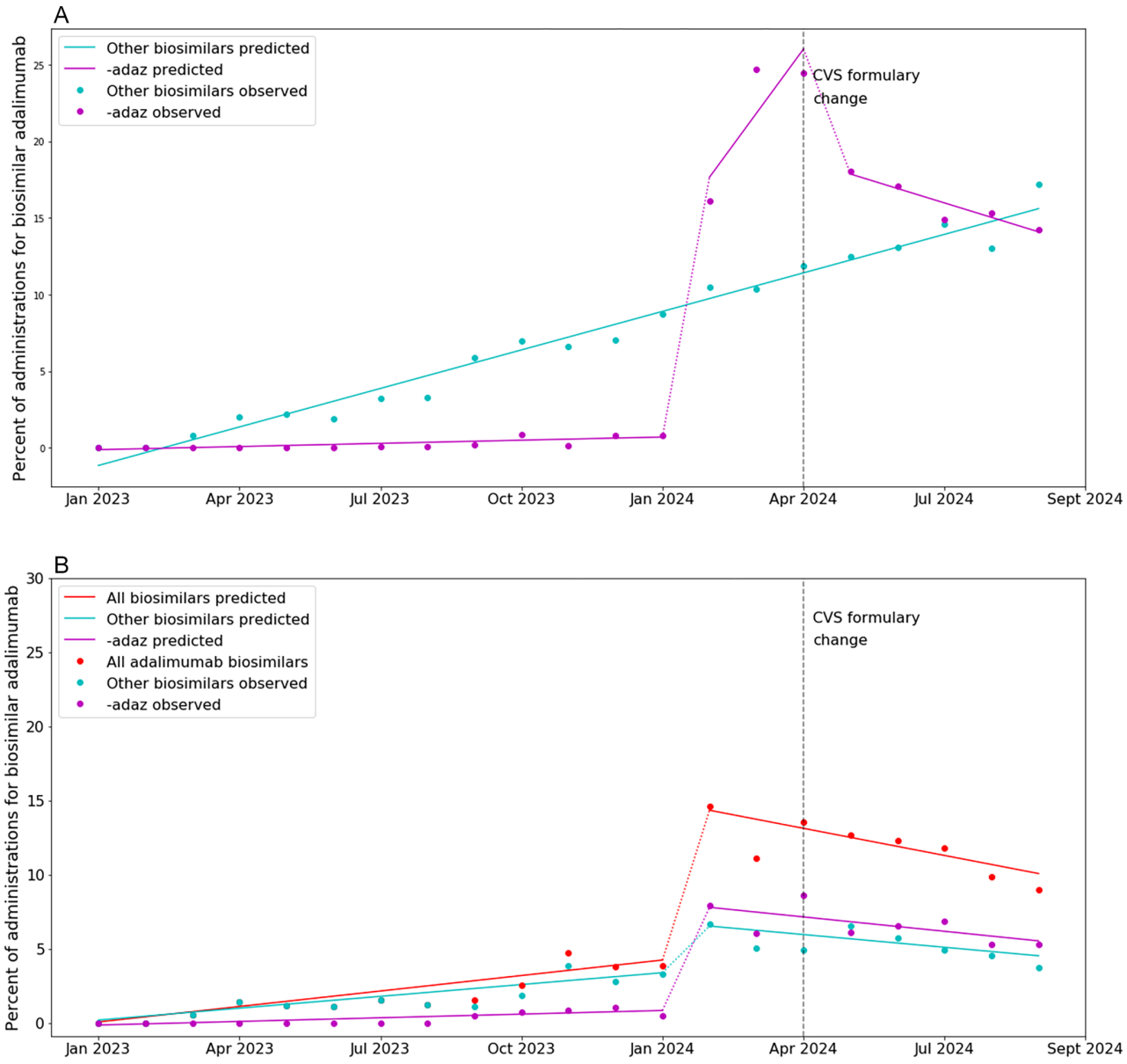

Figure 1. Interrupted time series analysis of the proportion of prescriptions written that were adalimumab-adaz and other adalimumab biosimilars by month stratified by insurance payor. (A) Biosimilar adalimumab uptake in privately insured patients. The models depicted are as follows:

$$P_{adalimumab-adaz} = \beta_0 + \beta_1 T + \beta_2 X_{14} + \beta_3 X T_{14} + \beta_4 X_{17} + \beta_5 X T_{17},$$

$$P_{other\ biosimilars} = \beta_0 + \beta_1 T.$$

(B) Biosimilar adalimumab uptake in Medicare patients. The models depicted are as follows:

$$P_{adalimumab-adaz} = \beta_0 + \beta_1 T + \beta_2 X_{14} + \beta_3 X T_{14},$$

$$P_{other\ biosimilars} = \beta_0 + \beta_1 T + \beta_2 X_{14} + \beta_3 X T_{14}.$$

Lines represent model-based predictions, and dots represent observed proportions. -adaz, adalimumab-adaz.

Medicare beneficiaries. Patients preferentially switched back to the form of adalimumab they started with, whereas new users of adalimumab-adaz were more likely to continue taking the biosimilar. This study is among the first to evaluate biosimilar dynamics in response to a national payer-driven formulary change and suggests that although formulary shifts can drive

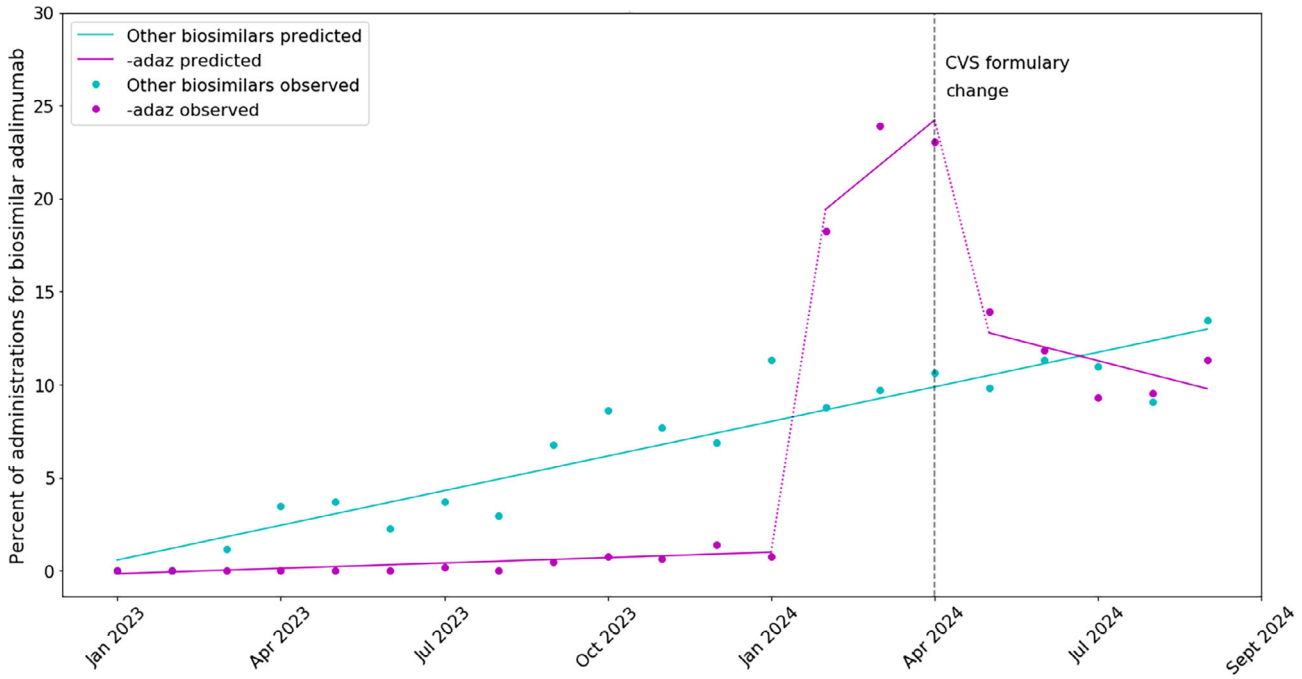

Figure 2. Interrupted time series analysis of the proportion of prescriptions written that were adalimumab-adaz and other adalimumab biosimilars by month among patients who are new users of adalimumab. Lines represent model-based predictions, and dots represent observed proportions. The models depicted are as follows:

$$P_{adalimumab-adaz} = \beta_0 + \beta_1 T + \beta_2 X_{14} + \beta_3 X T_{14} + \beta_4 X_{17} + \beta_5 X T_{17},$$

$$P_{other\ biosimilars} = \beta_0 + \beta_1 T.$$

-adaz, adalimumab-adaz.

rapid biosimilar uptake, they might be insufficient for durable biosimilar adoption.

Our findings are consistent with contemporaneous reports from the lay press.[25] We found that an increase in adalimumab-adaz prescriptions began approximately two months before the formulary change corresponding to the announcement of the policy, suggesting anticipatory prescribing behavior by clinicians and pharmacists. However, the observed and marked decline in prescribing volume shortly after the initial surge has not been previously reported. Product characteristics, such as a higher-concentration, lower-volume formulations, absence of citrate buffer, and use of a smaller-gauge needle, are known to influence patient experience with biologic drugs and have been cited as primary reasons for discontinuation or switching in prior studies.[36,37] However, these features are similar between adalimumab-adaz and the originator product, suggesting other factors, including patient price sensitivity or preference for the originator biologic, may play a more significant role in the observed decline.

Differences in copay assistance between originator and biosimilar products may contribute to decreased refill rates after the initial prescription for privately insured patients; copay assistance programs like drug coupons are prohibited for Medicare beneficiaries and not relevant to those with Medicaid. Although the bio-originator adalimumab offered up to $14,000 in copay assistance annually in 2025, branded biosimilars offered about one-third that amount. Differences in copay assistance may increase out-of-pocket costs for privately insured patients, potentially discouraging continued use. Patients are very sensitive to out-of-pocket costs,[38] so a large reduction in assistance is likely to have a negative effect on prescribing over the long term, but it is unclear if this change is driving large short-term reductions. Similarly, the removal of a patient support program and care supplies following a national switch has been found to negatively influence patients and providers.[19,39] In the United States, manufacturers advertise navigator programs directly to patients, but it is unclear how the introduction of biosimilars changed access to these in this case. Changes to patient copay and navigator programs represent huge barriers to adoption for patients. For practices, keeping up with the changing landscape of formularies, copay programs, and ancillary patient support programs represents a tremendous administrative burden, which may contribute to the high drop-off in prescriptions for a biosimilar and large number of patients who stopped using adalimumab-adaz, even if only temporarily.

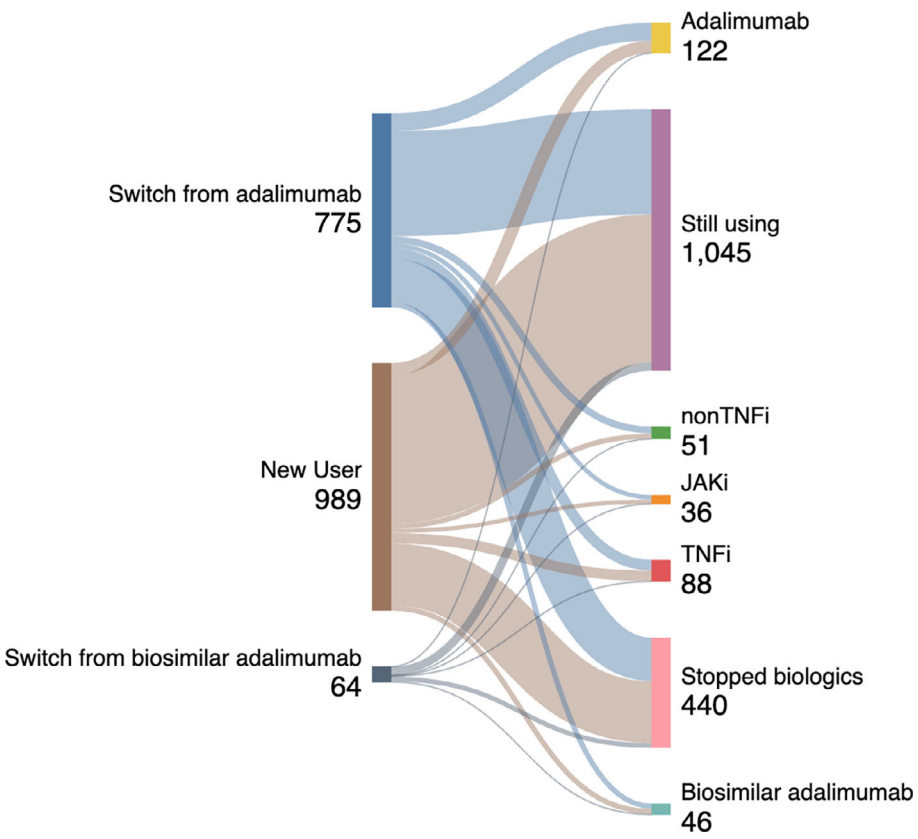

Figure 3. Pathways to and from use of adalimumab-adaz in 1,828 patients in the RISE between January 2023 and September 2024. A total of 27 patients with their first visit in the RISE corresponding to their index adalimumab-adaz prescription were omitted for parsimony because we could not reliably determine if they were a new user or switching. JAKi, JAK inhibitor; RISE, Rheumatology Informatics System for Effectiveness; TNFi, tumor necrosis factor inhibitor.

Our findings also suggest that the nocebo effect, a phenomenon whereby negative expectations regarding a treatment lead to the perception or experience of worsened symptoms, despite multiple studies documenting equivalent safety and effectiveness,[15,16] is contributing to the observed decrease in biosimilar use. This is likely a major driver of discontinuation among Medicare and Medicaid patients for whom copay assistance is not a factor. We observed that 19% of patients receiving adalimumab-adaz switched drugs, which is on par with that observed in open-label studies.[40] A recent review of strategies to mitigate the nocebo effect suggested a multimodal approach including incorporating patients and all health care providers (clinicians, nurses, and pharmacists), use of empathic, positive verbal and nonverbal communication, and providing information/trainings via multiple media at all stages of the process.[41]

Prescriptions for adalimumab-adaz increased among Medicare beneficiaries during the study, though to a lesser extent than commercially insured patients. Although the reasons for this are unclear, the trend may reflect prescribers streamlining their prescribing habits, increased clinician familiarity with adalimumab biosimilars, pharmacy stocking preferences, or formulary shifts within Medicare Part D plans. We also observed a statistically significant increase in prescribing for all other adalimumab biosimilar molecules in February 2024. An analysis of all Medicare Part D plans in January 2024 found nearly universal formulary coverage for originator adalimumab, whereas adalimumab-adaz had the second highest coverage among high-concentration formulas (about 27% of plans).[11] This study provided a snapshot of plan coverage in January 2024, so it is unclear when coverage changes took effect or any subsequent changes in coverage possibly in response to CVS's announcement, but this may explain the February 2024 jump. Critically though, prescriptions for biosimilar adalimumab are lower in Medicare than among patients with private insurance. When biosimilars are covered, formularies more frequently include the "high list price" version (the same biosimilar is offered at different list prices; high list price is ~5% lower than originator, whereas low list price is ~80% reduced), likely due to their appeal to PBMs, who profit off the rebate.[11,42]

The pattern of prescribing among new users, particularly the drop-off, was unexpected because clinicians report a greater willingness to prescribe biosimilars to new patients than to switch patients.[14,43,44] It has been reported that only one of three new patients prescribed an adalimumab biosimilar fill their prescription in the 30 days following their first attempt with rejections due to National Drug Code (NDC) blocks (30%), prior authorizations (14%), other rejections (14%), and abandonment primarily due to

high out-of-pocket costs (10%).[42] The drop-off in prescriptions may reflect clinicians adapting to insurance approval patterns. As they learn which prescriptions are likely to be denied, their prescribing behavior may shift accordingly. This trial-and-error process results in significant administrative burden for clinicians, nurses, and pharmacists and can lead to meaningful delays in starting or continuing treatment, which may be reflected in the large proportion of patients who stopped biologic therapy after using adalimumab-adaz.[39] The drop-off in prescriptions observed here, and resulting administrative burden and possible disruptions to care, is in stark contrast to what was observed in Canada following the implementation of new start and mandatory switching policies for etanercept and infliximab.[33]

Our analysis has strengths and limitations. First, the RISE collects medication reconciliation, e-prescriptions, and patient procedure data, which gives us comprehensive prescribing data. However, we did not have access to pharmacy fills.[45] It will be important for future work to replicate our findings with fill data and to compare e-prescriptions (representing orders) to fill data. Combined EHR and pharmacy fill data could help assess our hypotheses regarding prescription abandonment. Additionally, it is possible patients received adalimumab or one of its biosimilars from nonrheumatology sources and failed to report this during medication reconciliation; in a validation study comparing RISE data to Medicare claims, we found approximately 80% sensitivity and nearly perfect specificity for adalimumab.[45] Second, the RISE provides more contemporary data from multiple insurers compared to individual claims databases; however, we are unable to disaggregate insurance by subtype and therefore cannot assess whether the changes were specific to patients with a CVS Caremark formulary. Although we cannot directly assess the proportion of patients in the RISE whose pharmacy benefits are managed by CVS Caremark, the PBM holds substantial market share across most geographic markets in which RISE practices operate. Third, the RISE uniquely includes approximately 25% of community rheumatology practices nationwide, and emerging evidence shows practice type has an important influence on prescribing practices.[13,14,46] Additional work is needed to describe prescribing patterns in academic medical centers, 340B Drug Pricing Program–eligible organizations, and from other specialties that commonly prescribe adalimumab (eg, dermatology).

We observed significant increases in prescriptions for adalimumab-adaz for rheumatology patients nationwide following a major commercial formulary change. This is encouraging, particularly given extremely low uptake of biosimilar adalimumab in 2023. However, the subsequent decline in prescribing highlights ongoing barriers to sustained adoption. Future work to improve the long-term success of biosimilar rollouts should address several challenges. First, the nocebo effect will continue to plague biosimilar rollouts, underscoring the need for improved communication regarding biosimilars with patients. Second, prescribing biosimilars is associated with large administrative hurdles, including navigating copay assistance programs, addressing NDC blocks, prior authorizations, and rejections, disrupting patient care and placing a significant burden on clinicians, pharmacists, and nurses.[47] Third, the increases in prescriptions for adalimumab-adaz resulted from a commercial formulary change to specifically prefer a cobranded biosimilar that is part of a vertically integrated care chain rather than open market competition.[48] Although biosimilar entry has led to some price reductions,[11,49] market dynamics remain far from those envisioned by the Biologics Price Competition and Innovation Act (BPCIA). There are many billions in potential savings for patients, taxpayers, and employers that are still unrealized.[42] A comprehensive set of policy changes for the biosimilars market to bring it in line with the goals of the BPCIA is ambitious but closer than many people may realize.[5,50]

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 Roberts 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. IQVIA Institute. Biosimilars in the United States 2023-2027: competition, savings, and sustainability. IQVIA. January 31, 2023. Accessed February 15, 2024. https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/biosimilars-in-the-united-states-2023-2027
2. PR Newswire. Express scripts adds three biosimilars to largest formulary to promote competition and advance affordability, choices for patients. PR Newswire. July 10, 2023. Accessed April 2, 2025. https://www.prnewswire.com/news-releases/express-scripts-adds-three-biosimilars-to-largest-formulary-to-promote-competition-and-advance-affordability-choices-for-patients-301872877.html
3. Lee CC, Najafzadeh M, Kesselheim AS, et al. Cost to Medicare of delayed adalimumab biosimilar availability. Clin Pharmacol Ther 2021;110(4):1050–1056. doi:10.1002/cpt.2322
4. Zhai MZ, Sarpatwari A, Kesselheim AS. Why are biosimilars not living up to their promise in the US? AMA J Ethics 2019;21(8):E668–E678. doi:10.1001/amajethics.2019.668
5. Roberts ET, Yazdany J. Breaking barriers: essential reforms for a competitive and affordable biosimilars market. Curr Rheumatol Rep 2025;27(1):25. doi:10.1007/s11926-025-01192-x
6. Karas L, Sachs R, Anderson G. Legal Obstacles to Biosimilar Market Entry. Harv Public Health Rev (Camb) 2021;28. doi:10.54111/0001/bb2
7. Lexchin J. Affordable biologics for all. JAMA Netw Open 2020;3(4):e204753.
8. Mehr SR. Sustaining competition for biosimilars on the pharmacy benefit: use it or lose it. J Manag Care Spec Pharm 2024;30(6):600–603. doi:10.18553/jmcp.2024.30.6.600

9. Roberts ET, Schmajuk G, Yazdany J. Very low uptake of biosimilar adalimumab in the first 9 months of availability in rheumatology. Paper presented at: ACR Convergence 2024; November 17, 2024; Washington, DC. Accessed May 20, 2025. https://acrabstracts.org/abstract/very-low-uptake-of-biosimilar-adalimumab-in-the-first-9-months-of-availability-in-rheumatology/
10. Costin J, Mouslim MC, Socal MP, et al. Exploring the influence of health insurance plans on biosimilar adoption rates. PharmacoEconom Open 2024;8(1):115–118. doi:10.1007/s41669-023-00447-6
11. Klebanoff MJ, Li P, Lin JK, et al. Formulary coverage of brand-name adalimumab and biosimilars across Medicare part D plans. JAMA 2024;332(1):74–76. doi:10.1001/jama.2024.8917
12. Roberts ET, Li J, Bansback N, et al. Patterns of infliximab biosimilar uptake for Medicare, Medicaid, and private insurance from 2016-2022. Arthritis Rheumatol 2024;76(12):1739–1742. doi:10.1002/art.42963
13. Bond AM, Dean EB, Desai SM. The role of financial incentives in biosimilar uptake in Medicare: evidence from the 340B program. Health Aff (Millwood) 2023;42(5):632–641. doi:10.1377/hlthaff.2022.00812
14. Roberts ET, Bansback N, Tseng C-W, et al. Drivers of infliximab biosimilar uptake: a comparative analysis of new biosimilar initiations versus switching in a national rheumatology registry. Health Serv Res 2025;60:e14410.
15. Herndon TM, Ausin C, Brahme NN, et al. Safety outcomes when switching between biosimilars and reference biologics: a systematic review and meta-analysis. PLoS One 2023;18(10):e0292231. doi:10.1371/journal.pone.0292231
16. de Oliveira Ascef B, Almeida MO, de Medeiros-Ribeiro AC, et al. Therapeutic equivalence of biosimilar and reference biologic drugs in rheumatoid arthritis: a systematic review and meta-analysis. JAMA Netw Open 2023;6(5):e2315872.
17. Liu Y, Skup M, Yang M, et al. Discontinuation and switchback after non-medical switching from originator tumor necrosis factor alpha (TNF) inhibitors to biosimilars: a meta-analysis of real-world studies from 2012 to 2018. Adv Ther 2022;39(8):3711–3734. doi:10.1007/s12325-022-02173-7 Erratum in: Adv Ther 2022;39:5300.
18. Kakani P, Katcher BA, Maini L. Improved insurance coverage increased biosimilar Semglee's market share after the FDA's interchangeability designation. Health Aff (Millwood) 2025;44(11):1417–1425. doi:10.1377/hlthaff.2025.00443
19. Gasteiger C, Lobo M, Stanley R, et al. Rheumatology patients' experiences of a mandatory nationwide transition to an adalimumab biosimilar. ACR Open Rheumatol 2024;6(2):64–71. doi:10.1002/acr2.11634
20. Jhang J, Brennan TA. Evidence that regulatory and market forces are driving adoption of biosimilars. Health Aff (Millwood) 2024;43(11):1553–1560. doi:10.1377/hlthaff.2024.00366
21. Bates H, Lynch K, John N, et al. Pharmacy Benefit Manager (PBM) Overview. BRG Consulting; 2025.
22. National CooperativeRx. Pharmacy benefit management concepts: processes and strategies. National CooperativeRx. October 5, 2024. Accessed December 3, 2025. https://www.nationalcooperativerx.com/educational-materials/pharmacy-benefit-management-concepts-processes-and-strategies/
23. Fein AJ. The top pharmacy benefit managers of 2024: market share and key industry developments. Drug Channels. March 31, 2025. Accessed December 3, 2025. https://www.drugchannels.net/2025/03/the-top-pharmacy-benefit-managers-of.html
24. Volansky R. CVS Caremark to replace Humira with biosimilars on national commercial formularies. Healio. January 8, 2024. Accessed January 8, 2024. https://www.healio.com/news/rheumatology/20240105/cvs-caremark-to-replace-humira-with-biosimilars-on-national-commercial-formularies
25. Cohen JP. CVS Caremark's policy shift on Humira biosimilars may not be what the doctor ordered. Forbes. May 2, 2024. Accessed March 26, 2025. https://www.forbes.com/sites/joshuacohen/2024/05/02/cvs-caremarks-policy-shift-on-humira-biosimilars-may-not-be-what-the-doctor-ordered/
26. CVS Health. CVS Caremark accelerates biosimilars adoption through formulary changes. CVS Health. January 3, 2024. Accessed April 15, 2025. https://www.cvshealth.com/news/pbm/cvs-caremark-accelerates-biosimilars-adoption-through-formulary-changes.html
27. Yazdany J, Bansback N, Clowse M, et al. Rheumatology informatics system for effectiveness: a national informatics-enabled registry for quality improvement. Arthritis Care Res (Hoboken) 2016;68(12):1866–1873. doi:10.1002/acr.23089
28. Maroko AR, Doan TM, Arno PS, et al. Integrating social determinants of health with treatment and prevention: a new tool to assess local area deprivation. Prev Chronic Dis 2016;13:E128. doi:10.5888/pcd13.160221
29. 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;76:1037–1044.
30. Bayoumi AM. STDDIFF: Stata module to compute standardized differences for continuous and categorical variables. Statistical Software Components S458275. Boston College Department of Economics. Revised December 17, 2022. Accessed December 4, 2025. https://ideas.repec.org/c/boc/bocode/s458275.html
31. Faraone SV. Interpreting estimates of treatment effects: implications for managed care. P T 2008;33(12):700–711.
32. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. Stata J 2015;15(2):480–500. doi:10.1177/1536867X1501500208
33. McClean AR, Cheng L, Bansback N, et al. Uptake and spending on biosimilar infliximab and etanercept after new start and switching policies in Canada: an interrupted time series analysis. Arthritis Care Res (Hoboken) 2023;75(9):2011–2021. doi:10.1002/acr.25099
34. Schmajuk G, Ware A, Li J, et al. National rollout of a medication safety dashboard to improve testing for latent infections among biologic and targeted synthetic disease-modifying agent users within the Veterans Health Administration. Health Serv Res 2025;60(1):e14363. doi:10.1111/1475-6773.14363
35. Lopez Bernal J, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. J Clin Epidemiol 2018;103:82–91. doi:10.1016/j.jclinepi.2018.05.026
36. Rome BN, Kesselheim AS. Biosimilar competition for Humira is here: signs of hope despite early hiccups. Arthritis Rheumatol 2023;75(8):1325–1327. doi:10.1002/art.42520
37. Allegretti JR, Brady JH, Wicker A, et al. Relevance of adalimumab product attributes to patient experience in the biosimilar era: a narrative review. Adv Ther 2024;41(5):1775–1794. doi:10.1007/s12325-024-02818-9
38. Curkendall S, Patel V, Gleeson M, et al. Compliance with biologic therapies for rheumatoid arthritis: do patient out-of-pocket payments matter? Arthritis Rheum 2008;59(10):1519–1526. doi:10.1002/art.24114
39. Gasteiger C, Lobo M, Wong LS, et al. Health care providers' experiences of a mandatory nationwide transition to an adalimumab biosimilar. ACR Open Rheumatol 2023;5(12):644–651. doi:10.1002/acr2.11617
40. Odinet JS, Day CE, Cruz JL, et al. The biosimilar nocebo effect? A systematic review of double-blinded versus open-label studies. J Manag Care Spec Pharm 2018;24(10):952–959. doi:10.18553/jmcp.2018.24.10.952

41. Car E, Vandenplas Y, Lacosta TB, et al. Mitigating the nocebo effect in biosimilar use and switching: a systematic review. Pharmaceut Med 2024;38(6):429–455. doi:10.1007/s40290-024-00541-y
42. IQVIA. Adalimumab biosimilar tracking Q1 readout. SCRIBD. April 2, 2024. Accessed June 6, 2025. https://www.scribd.com/document/800126757/Humira-Biosimilar-Tracking-Q1-Readout-1712213405
43. Gibofsky A, McCabe D. US rheumatologists' beliefs and knowledge about biosimilars: a survey. Rheumatology (Oxford) 2021;60(2):896–901. doi:10.1093/rheumatology/keaa502
44. Cardinal Health. 2023 biosimilars report. 2023. Accessed October 15, 2024. https://www.cardinalhealth.com/content/dam/corp/web/documents/Report/cardinal-health-biosimilars-report-2023.pdf
45. Li J, Baker R, Stovall R, et al. Agreement of medication information derived from EHR data compared to Medicare insurance claims: an analysis of biologic disease-modifying antirheumatic drugs. Pharmacoepidemiol Drug Saf 2024;33(10):e70020. doi:10.1002/pds.70020
46. Dean EB, Johnson P, Bond AM. Physician, practice, and patient characteristics associated with biosimilar use in Medicare recipients. JAMA Netw Open 2021;4(1):e2034776.
47. Busis NA, Khokhar B, Callaghan BC. Streamlining prior authorization to improve care. JAMA Neurol 2024;81(1):5–6. doi:10.1001/jamaneurol.2023.4324
48. SmithRx. Unraveling PBM complexity: the consequences of vertical integration. SmithRx. March 28, 2024. Accessed December 20, 2024. https://smithrx.com/blog/unraveling-pbm-complexity-the-consequences-of-vertical-integration
49. Rome BN, Bhaskar A, Kesselheim AS. Use, spending, and prices of adalimumab following biosimilar competition. JAMA Health Forum 2024;5(12):e243964.
50. Sullivan SD, Humphreys SZ, Fox D, et al. Stakeholder perspectives on the sustainability of the United States biosimilars market. J Manag Care Spec Pharm 2024;30(10):1065–1072. doi:10.18553/jmcp.2024.24104

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1298–1303
DOI 10.1002/acr.80014
© 2026 American College of Rheumatology

BRIEF REPORT

Reproductive Health Counseling in Rheumatology: Gaps, Barriers, and Patient Impact

Daniele Marcy, MD,[1] Kristin Sturm, MA,[1] Tara Skorupa, MD,[2] Caroline Maretz, MD,[1] Guiset Carvajal Bedoya, MD,[3] Aaron Lazorwitz, MD, PhD,[4] Ju Young Kwag, BA,[1] M. Kristen Demoruelle, MD, PhD,[1] and JoAnn Zell, MD[1]

Objective. This study aimed to assess the frequency and factors associated with reproductive health counseling. We also aimed to assess barriers to counseling and introduce clinical interventions that could reduce these barriers.

Methods. A total of 101 female patients of childbearing potential who were not planning pregnancy completed a survey that included questions to determine whether they had previously received reproductive health counseling by a rheumatologist. This represented 18% of women aged 18 to 50 years seen in the rheumatology clinic. Four categories of clinical interventions were initiated with a goal to reduce barriers and improve effectiveness of counseling. Rheumatology providers were surveyed before and after implementing these interventions.

Results. Seventy percent of all patients and 85% of patients who took teratogenic medications reported having ever received reproductive health counseling by a rheumatologist. Counseling rates were higher in younger women and those who took teratogenic medications. Women who had previously received counseling were less likely to be unsure whether they were safe to become pregnant and more likely to talk to their rheumatologist if they decided to become pregnant in the future. Patient-reported rates of counseling poorly agreed with medical chart documentation of counseling, except in women taking teratogenic medications. Most providers found the clinical interventions made it easier to provide counseling, and the number of referrals to Family Planning clinic increased by 625%.

Conclusion. Reproductive health counseling for rheumatology patients of childbearing potential is essential. Ongoing efforts should focus on promoting consistent, high-quality counseling that translates into meaningful outcomes for patients.

INTRODUCTION

Reproductive health counseling is essential for rheumatology patients, as unplanned pregnancy is associated with poorer outcomes.[1] Given their ongoing relationships with their patients, rheumatologists are well-positioned to provide reproductive health counseling, including discussions of potential medication risks during pregnancy, supporting pregnancy planning for those who wish to conceive, and advising on appropriate contraceptive options for those who do not.

Yet, there remains a significant gap in the delivery of reproductive health counseling by rheumatologists. One prospective study of 206 women with systemic lupus erythematosus (SLE) found that 59% at risk of unintended pregnancy had not received contraception counseling within the past year.[2] A more recent cross-sectional study of 453 patients with SLE of childbearing age found higher rates of contraception counseling (78%) documented in the electronic medical record (EMR), but only 60% of that counseling came from rheumatologists.[3] In an observational study with 150 women of childbearing age with rheumatoid arthritis (RA), only 43% had EMR documentation of counseling, despite half taking methotrexate, a known teratogenic medication.[4]

Although there is evident opportunity to improve reproductive health counseling in rheumatology, rheumatologists face multiple barriers that limit its routine integration. These barriers include

Supported by the Ludeman Family Center for Women's Health Research at the University of Colorado Anschutz Medical Campus and funding from the NIH (grant T32-AR-07534).

[1]Division of Rheumatology, University of Colorado School of Medicine, Aurora; [2]Beth Israel Deaconess Medical Center, Division of Rheumatology, Boston, Massachusetts; [3]Department of Rheumatology, Billings Clinic, Billings, Montana; [4]Department of Obstetrics, Gynecology & Reproductive Sciences, Yale School of Medicine, New Haven, Connecticut.

Drs Demoruelle and Zell are co-senior authors and contributed equally to this work.

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80014.

Address correspondence via email to Daniele Marcy, MD, at Daniele.marcy@cuanschutz.edu.

Submitted for publication July 18, 2025; accepted in revised form January 22, 2026.

SIGNIFICANCE & INNOVATIONS

- Patients with underlying systemic autoimmune disease have higher rates of adverse pregnancy outcomes, yet 30% of patients surveyed in our study reported never receiving reproductive health counseling from their rheumatologist, including 16% of those on teratogenic medications.
- Patients who reported prior reproductive health counseling in our study were significantly less likely to feel unsure about pregnancy safety and more likely to discuss future pregnancy plans with their rheumatologist, highlighting the meaningful patient-centered impact of counseling.
- Implementing clinic tools aimed at improving the ease and efficacy of reproductive health counseling was associated with an increase in referrals for family planning services and made reproductive health counseling easier and more routine for rheumatology providers.

perceived lack of competency in the area, limited time when addressing other complex medical issues, preference for counseling to be provided by obstetricians or gynecologists (Ob/Gyn), and challenges in referring patients to Ob/Gyn.[5,6]

Our objectives in this study were to evaluate the frequency and factors associated with reproductive health counseling through a patient-reported survey and paired EMR review. We sought to identify perceived barriers to counseling among rheumatology providers and to implement clinical interventions that reduce these barriers and enhance reproductive health counseling by rheumatology providers.

PATIENTS AND METHODS

Study participants. An email invitation was sent to 556 female patients 18 to 50 years of age seen between September 2020 and March 2021 in the University of Colorado Hospital (UCH) outpatient rheumatology practice. We received responses from 276 patients (50%), who then completed a screening survey assessing additional eligibility criteria. The eligibility criteria included that women (1) were 18 to 50 years of age, (2) were diagnosed with any systemic autoimmune condition, (3) had no history of surgical sterilization or hysterectomy, (4) were English-speaking, and (5) had no plans for pregnancy in the next year, which was assessed by the question, "Do you wish to become pregnant in the next year?" Of the 276 respondents, 108 (50%) were not interested in participating, and 67 did not meet all eligibility criteria. In total, 101 met all inclusion criteria and were included in the study, which was 18% of the women aged 18 to 50 years seen in the clinic during those six months.

Participant questionnaire. Enrolled participants received an online survey that obtained self-report of demographics, including race and ethnicity from a fixed set of categories, current medications, and current use of contraception. To determine whether participants had received reproductive health counseling, they were asked, "Have you ever received counseling about contraception from your rheumatologist?" and "Have you ever been counseled by your rheumatologist about the risks of your medications if you become pregnant?" If participants reported having been counseled, they were asked "How well did you understand the counseling?" on a scale of 0 to 10, with 0 being "not at all" and 10 being "very well." Participants were also asked "Do you think you are safe to become pregnant now?" and "If you decided to become pregnant in the future, would you talk to your rheumatologist about those plans?"

Medical chart review of reproductive health counseling. Manual review of the EMR was performed by three co-authors (DM, TS, GCB) to identify provider documentation of reproductive health counseling. All rheumatology notes before enrollment were reviewed. For any chart where no counseling documentation was found, a second reviewer performed an independent chart review to confirm. Any discrepancies were adjudicated by a third reviewer.

Clinic interventions. To attempt to reduce barriers and improve the efficiency of reproductive health counseling, four clinical interventions were implemented in April 2021, after all participants completed their baseline study questionnaires: (1) *Check-in form*. A form was distributed by front desk staff to female patients <50 years of age at every clinic visit. The form was handed to their provider at the beginning of the appointment. The form asked about pregnancy plans using the question "Would you like to become pregnant in the next year?" and questions on current contraceptive use and interest in scheduling a Rheumatology Women's Clinic Visit. Before this intervention, there was no pathway for routinely assessing these questions. (2) *Rheumatology Women's Clinic visit*. This optional visit was dedicated to reproductive health counseling, including contraception and pregnancy planning. Counseling was provided by a rheumatology provider in the practice with clinical interest in the area. (3) *Patient education materials*. Educational handouts were placed in examination rooms to support provider-patient conversations. Materials included visual diagrams with information on contraception options and pregnancy planning considerations. (4) *EMR modifications*. The following changes were made to the EMR to support and prompt counseling: A pop-up box appeared when a teratogenic medication was ordered with a warning of teratogenicity and a direct link to a referral order for our affiliated Family Planning Clinic, where initiation of contraception can be facilitated. Additionally, an Epic smart phrase and note template

were developed to streamline documentation of reproductive health counseling.

Provider questionnaire. Two anonymous voluntary online questionnaires were distributed to clinic providers to assess how often they provide reproductive health counseling and perceived barriers. The first was sent before, and the second one year after, implementation of the above-mentioned clinical interventions. The second survey was sent only to providers who completed the first, to assess change in practices or perceptions. The initial survey was completed by 20 of 22 (91%) and the second by 14 of 20 (70%) providers (8 rheumatology attending physicians, 5 rheumatology fellows, and 1 rheumatology advance practice provider).

Statistics. Factors associated with a self-reported history of counseling were compared using chi-square or Wilcoxon rank-sum test, as appropriate. The agreement between patient-reported counseling and provider documentation of counseling was analyzed using Cohen's κ (where k = 1 represents perfect agreement and k = 0 represents no agreement). The number of referrals sent to our affiliated Family Planning clinic were compared before and after the clinical interventions using Wilcoxon signed-rank test.

Study approval. Study procedures were approved by the Colorado Multiple Institutional Review Board and in compliance with the Declaration of Helsinki. Written informed consent was obtained from all patients.

RESULTS

Participant characteristics are included in Table 1. The majority of patients had a diagnosis of SLE (33%) or RA (33%). Thirty-nine (39%) were currently or had previously taken medications with known teratogenic effects, and 32 (32%) were currently taking teratogenic medications.

Participant-reported rates of reproductive health counseling. Sixty-seven (67%) women reported having ever been counseled about the potential risks of their medications if they were to become pregnant, 41 (41%) reported having ever been counseled about contraception, and 70 (70%) reported having had at least one of these two forms of reproductive health counseling by their rheumatologist. Of the 39 women currently or previously taking teratogenic medications, 33 (84.6%) reported having ever been counseled about the potential teratogenic risks of their medications, and 23 (59.0%) had ever been counseled about contraception by their rheumatologist.

Reported rates of counseling were higher in women who took teratogenic medications ($P = 0.01$) and were aged <40 years ($P < 0.01$), and there was a trend toward higher counseling rates in patients with SLE ($P = 0.08$) (Table 2).

Table 1. Patient characteristics*

Characteristics	n (%)
Rheumatologic diagnosis	
RA	33 (33)[a]
SLE	33 (33)[a]
Other systemic autoimmune disease	37 (37)
Age	
<40 y	75 (74)
≥40 y	26 (26)
Race and ethnicity[b]	
Asian	4 (4)
Black or African American	1 (1)
Hispanic or Latino/a	21 (21)
Multiracial	10 (10)
Not Hispanic or Latino/a	77 (77)
Other race	6 (6)
White	79 (79)
Medications currently taking	
Methotrexate	16 (16)
Leflunomide	1 (1)
Cyclophosphamide	0 (0)
Mycophenolate	13 (13)
Coumadin	4 (4)
Any medication with established teratogenic risk[c]	32 (32)
Non-TNF biologic or tsDMARD	19 (19)
Any medication recommended to discontinue in pregnancy[d]	46 (46)
Medications currently or previously taking	
Methotrexate	19 (19)
Leflunomide	2 (2)
Cyclophosphamide	1 (1)
Mycophenolate	17 (17)
Coumadin	4 (4)
Any medication with established teratogenic risk[c]	39 (39)
Non-TNF biologic or tsDMARD	23 (23)
Any medication recommended to discontinue in pregnancy[d]	53 (53)
Education[e]	
High school or less	7 (7)
Some college	21 (21)
College graduate	38 (38)
Postdoctoral	33 (33)
Pregnancy history	
Ever pregnant	50 (50)
Never pregnant	51 (50)
Marital status[e]	
Married	40 (40)
Living with someone/registered partnership	14 (14)
Widowed/divorced/separated	8 (8)
Single	37 (37)
Current contraceptive method[f]	
None	14 (14)
Less effective	7 (7)
Effective	11 (11)
Highly effective	68 (68)

* RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; TNF, tumor necrosis factor; tsDMARD, targeted synthetic disease-modifying antirheumatic drug.
[a] Two participants had an RA/SLE overlap.
[b] Two participants chose not to answer.
[c] Medication with established teratogenic effects: methotrexate, leflunomide, mycophenolate, cyclophosphamide, and coumadin.
[d] Medications recommended to be discontinued during pregnancy: teratogenic medications, JAK inhibitors, and any biologic other than anti-TNF inhibitors.
[e] Two participants chose not to answer.
[f] Effectiveness of contraception defined as (1) highly effective: intrauterine device, hormonal implant, sterilization of male partner, only female partners, abstinence; (2) effective: oral contraceptive pills, vaginal ring, or hormonal injection; (3) less effective: condoms or withdrawal method. One participant chose not to answer.

Table 2. Factors influencing whether patients reported having ever received reproductive health counseling*

	Women reporting any counseling,[a] %	*P* value
Type of autoimmune disease		
SLE (n = 33)	82	
Other (n = 68)	65	0.08
Teratogenic medication use		
Ever (n = 39)	85	
Never (n = 62)	61	**0.01**
Teratogenic medication use		
Current (n = 32)	84	
Not currently (n = 69)	64	**0.04**
Self-reported race/ethnicity		
Non-Hispanic White (n = 64)	65	
Other (n = 35)	35	0.12
Marital status		
Married (n = 59)	60	
Other (n = 40)	40	0.62
Age		
≥40 y (n = 26)	50	
<40 y (n = 75)	77	**<0.01**
Pregnancy history		
Ever pregnant (n = 50)	64	
Never pregnant (n = 51)	76	0.17
Education level		
Completed college (n = 71)	70	
Did not complete college (n = 28)	68	0.80

* Bold indicates significance. SLE, systemic lupus erythematosus.
[a] Women reporting that they had ever previously received counseling on contraception or the risks of their medications if they were to become pregnant.

Effectiveness of reproductive health counseling. Women who had received counseling reported good understanding of the counseling (median [interquartile range] on a 10-point scale; understanding for medication counseling: 10 [9–10]; understanding of contraception counseling: 10 [9.75–10]). Notably, women who reported having received reproductive health counseling were less likely to answer "unsure" when asked whether they were safe to become pregnant, compared to women who did not report prior counseling (9% vs 37%; $P < 0.01$). This difference was also present in the subset of women currently taking teratogenic medications (0% vs 40%; $P = 0.02$). Women who had previously received counseling from their rheumatologist were also more likely to report that they would talk to their rheumatologist if they decided to become pregnant in the future (99% vs 83%; $P < 0.01$); a difference also present in the subset of women taking teratogenic medications (100% vs 80%; $P = 0.02$). Women reporting prior reproductive health counseling were not more likely to use highly effective or effective forms of contraception compared to women who had not received prior counseling (79% vs 80%; $P = 0.87$).

Self-reported counseling compared to medical chart documentation. Overall, agreement was only fair between women reporting having received reproductive health counseling and women with chart documentation of counseling (k = 0.27, Table 3). In women self-reporting a history of counseling, 75% had EMR documentation, and in women self-reporting no history of counseling, 47% had EMR documentation of counseling. However, agreement was higher for women currently taking teratogenic medications (k = 0.59).

Provider-reported rates of contraception counseling and reported barriers. For providers who completed pre- and postintervention surveys, before the clinical interventions initiated in this study, 7 of 14 (50%) reported providing reproductive health counseling at most clinic visits for patients of childbearing potential, and 7 (50%) reported counseling less frequently (eg, annually, less than annually, at their initial visit, if they changed immunosuppression, if the patient brings it up, or not at all). The most frequently cited barriers to providing counseling were too many other active medical issues (86%) and short length of clinic visits (79%). After introduction of our clinical interventions, the number of providers providing reproductive health counseling at most or every clinic visit for patients of childbearing potential increased from 7 of 14 (50%) to 10 of 14 (71%). Nine of 14 (64%) reported that the clinical changes made it easier to counsel patients about contraception or pregnancy planning. Changes most reported by providers as facilitating reproductive health counseling were the check-in form (71%) and option for a rheumatology women's clinic visit (50%).

Changes in referrals to Family Planning clinic. In the two years before our clinical interventions (April 2019 to March 2021), the total number of clinical referrals from rheumatology providers to the UCH Ob/Gyn Family Planning clinic was 9 (0.4 per month). In the 12 months after the clinical interventions (April 2021 to March 2022), referrals increased to 35 (2.9 per month; $P = 0.02$).

DISCUSSION

Reproductive health counseling is an important component of rheumatologic care that can meaningfully improve patient

Table 3. Agreement between patient-reported reproductive health counseling and medical chart documentation of reproductive health counseling*

	Agreement[a]
All women	0.27
Women ever on a teratogenic medication	0.43
Women currently on a teratogenic medication	0.59
Women with SLE	0.31
Women with RA	0.39

* RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.
[a] Agreement calculated based on Cohen's κ statistic, with 0.01–0.20 = slight agreement, 0.21–0.40 = fair agreement, 0.41–0.60 = moderate agreement, 0.61–0.8 = substantial agreement, 0.81–1 = almost perfect agreement.

outcomes and clinical experiences.[7] Patients report a preference for their rheumatologist to proactively initiate and revisit reproductive health conversations, provide clear information about pregnancy and fertility risks, and address reproductive health in the context of their personal lives and values.[8]

In this study, we found that a substantial proportion (70%) of patients of childbearing potential had previously received reproductive health counseling by a rheumatologist. Yet, there remained 30% of women, including 16% currently taking teratogenic medications, who reported never having been counseled. Although these rates are comparable or higher than those found in other studies,[3,9,10] they still underscore a need for improvement and proactive rheumatologist engagement.

We found that rates of counseling were higher in younger women and women taking teratogenic medications. Although this might be expected, it also highlights areas in which counseling rates could be improved. Older women and those not on teratogenic medications may also have significant reproductive health risks and should also receive counseling. It is encouraging that we found high rates of counseling in women with lupus, given their established risk for pregnancy complications.[11] Yet, women with RA and other autoimmune diseases also face higher risks, particularly those with active disease at conception. Reproductive health counseling should remain a priority across all patients with rheumatologic diseases.

Evaluating the true impact of reproductive health counseling can be challenging. There is often a disconnect between reported counseling, either by EMR documentation or patient report, and evidence of behavioral change or improved reproductive health outcomes. An observational study of 453 reproductive-aged women with SLE found that 78% of patients had EMR documentation of counseling, yet only 37% were using highly effective contraception.[3] Although our study also did not find prior counseling to be associated with increased use of effective contraception, we did find that prior counseling was linked to other meaningful indicators of effectiveness. We found that women who had previously received counseling were more confident in understanding pregnancy safety and more likely to initiate discussions with their rheumatologist about future pregnancy plans. Other studies have also reported higher patient comfort to discuss reproductive health issues with their rheumatologists when these topics were routinely discussed.[8] Notably, not all studies have found a relationship between counseling and patient confidence of pregnancy risk,[12] highlighting an important area for future research to identify the specific elements in provider-patient interactions that contribute to the most effective counseling.

In this study, we found poor agreement between patient-reported rates of reproductive health counseling and documentation in the EMR. However, this agreement was higher in women taking teratogenic medications. This finding should be considered when interpreting studies of counseling that rely only on EMR documentation.

In recent years, several studies have focused on ways to increase rates of reproductive health counseling and documentation, but with variable success.[6,13] We identified several interventions that facilitated counseling by rheumatology providers, including a check-in form given to patients at each visit that asked about their pregnancy plans and an option for a Rheumatology Women's Health visit that was fully dedicated to reproductive health counseling. The implementation of these strategies resulted in a remarkable 625% increase in referrals to the Ob/Gyn Family Planning Clinic. Such referrals are a critical component of effective counseling, as they can enable access to contraception for patients not planning pregnancy.

Strengths of our study include the inclusion of all women with rheumatologic disease, unlike many prior studies that focus exclusively on patients with SLE[4,5,12] or RA.[6] Additionally, by comparing patient self-report to EMR documentation, we identified a meaningful discrepancy between perceived and recorded counseling, which has not clearly been addressed in prior studies. Limitations of our study include the inability to assess the effectiveness of individual clinical interventions and lack of data on the timing and frequency of prior counseling. Additionally, we did not collect patient feedback or satisfaction with new tools or interventions. Lastly, our study population was well educated and frequent users of highly effective contraception, which may limit generalizability. These gaps represent important areas for future research.

In conclusion, our study underscores the importance of reproductive health counseling for women with rheumatologic diseases and persistent gaps in its delivery. Although many patients in our study reported having received counseling, many had not. Our findings highlight key disparities in who receives counseling and the value of targeted clinical interventions to support providers. As the field continues to recognize the central role rheumatologists play in reproductive health, future efforts should prioritize systemwide implementation of effective counseling strategies and integration of patient perspectives to ensure that all individuals receive timely, person-centered reproductive health care.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Marcy 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. Rajendran A, Eudy AM, Balevic SJ, et al. The importance of pregnancy planning in lupus pregnancies. Lupus 2021;30(5):741–751.

2. Yazdany J, Trupin L, Kaiser R, et al. Contraceptive counseling and use among women with systemic lupus erythematosus: a gap in health care quality? Arthritis Care Res (Hoboken) 2011;63(3):358–365.
3. Silverstein RG, Fitz V, Thornton M, et al. Contraceptive use and counseling in patients with systemic lupus erythematosus. Contraception 2022;105:46–50.
4. Ingram E, Claus L, Kolfenbach J, et al. Contraceptive use in women of childbearing ability with rheumatoid arthritis. J Clin Rheumatol 2021; 27(3):97–101.
5. Murray S, Augustyniak M, Murase JE, et al. Barriers to shared decision-making with women of reproductive age affected by a chronic inflammatory disease: a mixed-methods needs assessment of dermatologists and rheumatologists. BMJ Open 2021;11(6): e043960.
6. Pryor KP, Albert B, Desai S, et al. Pregnancy intention screening in patients with systemic rheumatic diseases: pilot testing a standardized assessment tool. ACR Open Rheumatol 2022;4(8):682–688.
7. Sammaritano LR, Bermas BL, Chakravarty EE, et al. 2020 American College of Rheumatology guideline for the management of reproductive health in rheumatic and musculoskeletal diseases. Arthritis Rheumatol 2020;72(4):529–556.
8. Wolgemuth T, Stransky OM, Chodoff A, et al. Exploring the preferences of women regarding sexual and reproductive health care in the context of rheumatology: a qualitative study. Arthritis Care Res (Hoboken) 2021;73(8):1194–1200.
9. Birru Talabi M, Clowse MEB, Schwarz EB, et al. Family planning counseling for women with rheumatic diseases. Arthritis Care Res (Hoboken) 2018;70(2):169–174.
10. Buasawat P, Manonai J, Ngamjanyaporn P. Perception of contraceptive counseling and contraceptive use among systemic lupus erythematosus patients. Contraception 2021;103(1):38–43.
11. Buyon JP, Kim MY, Guerra MM, et al. Predictors of pregnancy outcomes in patients with lupus: a cohort study. Ann Intern Med 2015; 163(3):153–163.
12. Brites L, Silva S, Andreoli L, et al. Effectiveness of reproductive health counseling of women with systemic lupus erythematosus: observational cross-sectional study at an academic lupus clinic. Rheumatol Int 2021;41(2):403–408.
13. Wolfgang T, Anstett S, Arabelovic S. Improvement of pregnancy counselling and contraception counselling and documentation in a single rheumatology academic practice: a quality improvement project. BMJ Open Qual 2022;11(4):e001871.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 78, No. 9, September 2026, pp 1304–1305
© 2026 American College of Rheumatology

LETTER

DOI 10.1002/acr.80041

Within-day activity patterns in knee osteoarthritis: frailty signal or treatment marker? Comment on the article by Liles et al

To the Editor:

Liles et al recently published a study titled "The relation of within-day physical activity patterns with all-cause mortality in adults with knee osteoarthritis: findings from the Osteoarthritis Initiative."[1] In this article, the authors report that a pattern of low activity in both the morning and evening is associated with a two-fold higher risk of death and propose within-day activity patterns as potential clinical markers to guide treatment decisions in knee osteoarthritis (OA).

In our view, an important conceptual issue deserves clarification: to what extent do these within-day patterns truly capture OA-specific treatment need, as opposed to broader disruption of rest–activity rhythms and frailty that happen to be common in people with OA? Work in community-dwelling older adults shows that fragmented or blunted 24-hour activity rhythms and specific diurnal activity profiles are associated with mortality and incident major disease even in populations unselected for OA.[2–4] Recent analyses from the National Health and Nutrition Examination Survey, for example, found that an "early-morning" physical activity pattern and an irregular sleep pattern were each independently related to higher all-cause mortality after adjustment for total activity volume and sleep duration and were strongly patterned by age, cardiometabolic comorbidity, and socioeconomic factors.[4] These findings suggest that when activity happens may be a marker of global vulnerability rather than disease-specific symptom control.

The same concern may apply here. The functional principal component analysis is restricted to a waking window and models covariates such as age and comorbidity, but the resulting components still likely reflect composite rest–activity rhythms influenced by sleep timing, work and caregiving schedules, depressive symptoms, and multimorbidity. Prior studies using hourly accelerometry have shown that frailer older adults accumulate less activity throughout the day, particularly in the morning and late evening, even after accounting for chronic conditions.[3] If the "low morning and evening" pattern in the Osteoarthritis Initiative largely indexes this broader frailty–circadian construct, then its association with mortality may tell us relatively little about knee-specific symptom burden or the need to escalate OA therapy.

This has practical implications. Framing within-day activity profiles as OA treatment markers could unintentionally shift attention away from modifiable determinants of rest–activity disruption—sleep disorders, social isolation, polypharmacy, or deconditioning—that might be more appropriate intervention targets. Conversely, if these patterns are interpreted as integrated markers of overall physiologic and social reserve in a cohort enriched for OA, they could complement rather than replace conventional OA assessments. Linking activity patterns with measures of frailty, cognitive function, depression, and 24-hour rest–activity rhythm metrics and testing whether OA-directed treatments actually modify these profiles would help disentangle OA-specific signals from global aging signals.

We therefore suggest that the authors' findings may be better understood as evidence that daily activity timing encodes information about whole-person vulnerability in people with or at risk for knee OA rather than as a direct indicator of when OA-specific treatment should be intensified. Clarifying this distinction would, in our view, avoid overinterpreting a promising prognostic signal as a disease-specific biomarker and would align their work with the broader literature on rest–activity rhythms, frailty, and mortality.

Drs Cheng and He are co-first authors and contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.80041.

Ruoyao Cheng, MB
ruoyaocheng136@outlook.com
Caihua He, MB
First Clinical Medical College
Gansu University of Chinese Medicine
Lanzhou, China
Ying Liu, MB
First Clinical Medical College
Gansu University of Chinese Medicine
Lanzhou, China
Clinical College of Traditional Chinese Medicine
Gansu University of Chinese Medicine
Lanzhou, China

1. Liles SC, Cleveland RJ, Jakiela JT, et al. The relation of within-day physical activity patterns with all-cause mortality in adults with knee osteoarthritis: findings from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2026;78(8):1058–1068. doi:10.1002/acr.70024
2. de Feijter M, Lysen TS, Luik AI. 24-h activity rhythms and health in older adults. Curr Sleep Med Rep 2020;6(2):76–83. doi:10.1007/s40675-020-00170-2
3. Huisingh-Scheetz M, Wroblewski K, Kocherginsky M, et al. The relationship between physical activity and frailty among U.S. older adults based on hourly accelerometry data. J Gerontol A Biol Sci Med Sci 2018;73(5):622–629. doi:10.1093/gerona/glx208
4. Zhang Y, Kivimäki M, Carrillo-Larco RM, et al. Diurnal patterns of accelerometer-measured physical activity and sleep and risk of all-cause mortality: a follow-up of the National Health and Nutrition Examination Surveys (NHANES). Int J Behav Nutr Phys Act 2024;21(1):120. doi:10.1186/s12966-024-01673-9

DOI 10.1002/acr.80042

Reply

To the Editor:

We read with great enthusiasm the letter by Cheng et al about our study.[1] In particular, we believe that an important question is asked, which is whether osteoarthritis (OA) symptoms are related to within-day patterns. The authors posit that frailty may disrupt within-day patterns and then cite that data from the National Health and Nutrition Examination Survey show that within-day patterns identified using K-means clustering algorithms are associated with mortality from the general population.[2] This article also assessed factors associated with such patterns. Cheng et al suggest that the within-day patterns may include "information about whole-person vulnerability" in adults with or at risk of knee OA.

Simply, we agree. We did not investigate the reasons why some had more representation of certain patterns than others. That was beyond the scope of our study. However, we do feel there is utility in employing within-day patterns as a vital sign or indicator of overall health. This concept is not new. Golightly and colleagues investigated measures of physical activity as a vital sign though a systematic review of the literature.[3] Future research should investigate if the emergence of particular within-day patterns is associated with the onset or worsening of symptoms, which may then further pave the way for clinicians to understand what aspects of physical activity may serve as a vital sign of health.

Daniel K. White, PT, ScD, MSc
dwk@udel.edu
Sydney C. Liles, PT, DPT, MS
Department of Physical Therapy
University of Delaware
Newark

1. Liles SC, Cleveland RJ, Jakiela JT, et al. The relation of within-day physical activity patterns with all-cause mortality in adults with knee osteoarthritis: findings from the Osteoarthritis Initiative. Arthritis Care Res (Hoboken) 2026;78(8):1058–1068. doi:https://doi.org/10.1002/acr.70024
2. Zhang Y, Kivimäki M, Carrillo-Larco RM, et al. Diurnal patterns of accelerometer-measured physical activity and sleep and risk of all-cause mortality: a follow-up of the National Health and Nutrition Examination Surveys (NHANES). Int J Behav Nutr Phys Act 2024;21(1):120. doi:https://doi.org/10.1186/s12966-024-01673-9
3. Golightly YM, Allen KD, Ambrose KR, et al. Physical activity as a vital sign: a systematic review. Prev Chronic Dis 2017;14:170030. doi: https://doi.org/10.5888/pcd14.170030