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 77 No. 5 May 2025

Arthritis Care & Research

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

Cover design: Sandra Pulmano

♾ This journal is printed on acid-free paper.

Arthritis Care & Research

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

VOLUME 77 • MAY 2025 • NO. 5

Cover image: The image on the cover (Prakash et al; pages 564–572) shows magnetic resonance images of the patient's brain at admission, indicating axial fluid-attenuated inversion recovery (left) and axial T2 sequences (center) showing large hyperintense lesions in right parietal, left frontoparietal, and left parietotemporal subcortical and deep white matter regions with increase in lesion size and perilesional edema; and new-onset, irregular contrast (gadolinium) enhancement with moderate peri-lesional edema and no mass effect (right). Magnetic resonance spectroscopy (MRS) indicates increased choline peak and modestly elevated lipid/lactate peaks (bottom).

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 557–563
DOI 10.1002/acr.25483

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Advancing Rheumatologic Care in Older Adults: Highlights From the 2024 American Geriatrics Society Annual Scientific Meeting

Jiha Lee,[1] Sarah B. Lieber,[2] Sebastian E. Sattui,[3] Namrata Singh,[4] Katherine D. Wysham,[5] and Una E. Makris[6]

Rheumatologists are caring for a rapidly increasing older adult population. Aging introduces certain complexities in rheumatology care, including the following: accumulation of comorbid conditions influences treatment choices and response, polypharmacy increases the risk and burden of treatments, and late-onset rheumatic diseases can have atypical presentations.[1,2] Patient preferences and goals of care, which often change over an individual's lifespan, impact therapeutic decision-making.

The American Geriatrics Society (AGS) stands at the forefront of addressing best practices for the care of older adults and can serve as an important partner to the American College of Rheumatology (ACR) and the rheumatology community at large. The AGS, founded in 1942, is a nationwide, not-for-profit society of geriatrics health care professionals dedicated to improving the health, independence, and quality of life of older people. In response to the growing population of older adults, the AGS has set out to provide support in education, training, dissemination, and implementation of geriatric care principles in medical and surgical specialties.[3] Several medical and surgical societies, including those representing emergency medicine,[4] oncology,[5,6] and cardiology[7–10] have embraced the integration of geriatric principles[11] and successfully developed national clinical and research networks to advance the care of older adults.[4] Given that the aging population will require rheumatology care, the need for comprehensive training in aging principles for rheumatologists has never been more urgent.

In this article, we provide an update from the 2024 AGS Annual Scientific Meeting (AGS 2024) that emphasizes key geriatric principles and concepts pertinent to rheumatologists. The coauthors believe that these sessions highlight opportunities to enhance clinical care for older adults with rheumatic diseases. Overall, the aim of this article is to raise awareness and disseminate knowledge gained from AGS 2024 to encourage dialogue on integrating comprehensive age-friendly rheumatologic care.

Approach

For this article, we selected material of highest relevance to the rheumatology community, including content related to clinical care, education, and research and anchored it on the Geriatrics 5Ms framework (Figure 1).[12] The Geriatric 5Ms, developed and endorsed by the AGS, Institute for Healthcare Improvement, and the John A. Hartford Foundation, supports the goal of delivering age-friendly health care by emphasizing mind, mobility, medications, what matters most, and multicomplexity.[12,13] We also highlight potential barriers to adopting aging-friendly practices in rheumatology. This summary concludes with a case study on how another specialty has integrated many of these principles into practice.

Dr Lee's work was supported by the National Institute on Aging (NIA), NIH (grant K23-AG-082727). Dr Lieber's work was supported by the NIA, NIH (grants K23-AG-088502 and R03-AG-082927). Dr Sattui's work was supported by Bristol Myers Squibb Foundation Robert A. Winn Diversity in Clinical Trials Career Development Award; Rheumatology Research Foundation Investigator Award; and NIA, NIH (grant R03-AG-082983). Dr Singh's work was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant K23-AR-079588); and NIA, NIH (grant R03-AG-082857). Dr Wysham's work was supported by the Department of Veterans Affairs (award CX002351). Dr Makris's work was supported in part by Veterans Affairs Health Systems Research (grants IIR 20-256 and HX003350); and NIA, NIH (grant P30-AG-022845).

[1]Jiha Lee, MD, MHS: University of Michigan, Ann Arbor; [2]Sarah B. Lieber, MD, MS: Hospital for Special Surgery and Weill Cornell Medicine, New York, New York; [3]Sebastian E. Sattui, MD, MS: University of Pittsburgh, Pittsburgh, Pennsylvania; [4]Namrata Singh, MD, MSCI: University of Washington, Seattle; [5]Katherine D. Wysham, MD: University of Washington and Veterans Affairs Puget Sound Health Care System, Seattle; [6]Una E. Makris, MD, MSc: University of Texas Southwestern Medical Center and Veterans Affairs North Texas Health Care System, Dallas.

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25483.

Address correspondence via email to Una E. Makris, MD, MSc, at una.makris@utsouthwestern.edu.

Submitted for publication November 26, 2024; accepted in revised form November 27, 2024.

SIGNIFICANCE & INNOVATIONS

- Incorporating geriatric principles, along with pursuing collaborative approaches with geriatricians and allied health professionals, can improve the care of older adults with rheumatic diseases.
- Assessing falls and social isolation, in addition to traditional rheumatologic approaches focused on joint mobility and function, can enhance shared decision-making.
- Applying the 5Ts framework (target population, team, time, tips to accommodate, tools) in research can enhance the inclusion of older adults in rheumatology research, which is essential for developing evidence-based age-friendly care.
- Adopting innovative, patient-centered care models that include comprehensive geriatric assessments focused on what matters most to the patient will help transform rheumatology into a more age-friendly specialty, in alignment with recent health care policy developments and priorities.

Clinical highlights: practical strategies incorporating a biopsychosocial approach to assessing and managing what matters most to older adults

Fall assessment and prevention. In older adults, falls are associated with an increased risk of fracture, disability, and mortality.[14] Individuals with rheumatic diseases may have an even higher fall risk because of joint pain and deformities, deconditioning, and medications.[15] Given the impact on both patient outcomes and health care costs, geriatricians consider falls a chronic health condition requiring focused management similar to any other chronic disease state.[16] Falls and instability are underreported by patients both because of under recognition, especially near misses in which injuries do not occur, as well as stigma (ie, concern that reporting falls may lead to clinical interventions that limit independence).[17]

In a session led by Elizabeth Eckstrom, MD, MPH, and entitled "Safe Mobility: The Role of Falls Assessment and Prevention in Clinical Practice," presenters suggested that fall assessment and prevention should be communicated to patients as "safe mobility," with an emphasis on reaching mobility goals and preserving independence.[18] The speakers in this session discussed the implementation of fall prevention programs in primary care settings and telemedicine visits using the CDC's Stopping Elderly Accidents, Deaths & Injuries (STEADI) Centers for Medicare and Medicaid Services (CMS) framework: screening (patients at risk), assessing (modifiable risk factors), and intervention (to reduce identified risk factors).[19] This framework can be implemented readily in clinical practice with three brief questions: (1) Do you feel unsteady when standing or walking?; (2) Do you have any worries about falling?; and (3) Have you fallen in the past year? A "yes" to any of those questions highlights the need for fall assessment and introduction of intervention strategies. We believe this simple

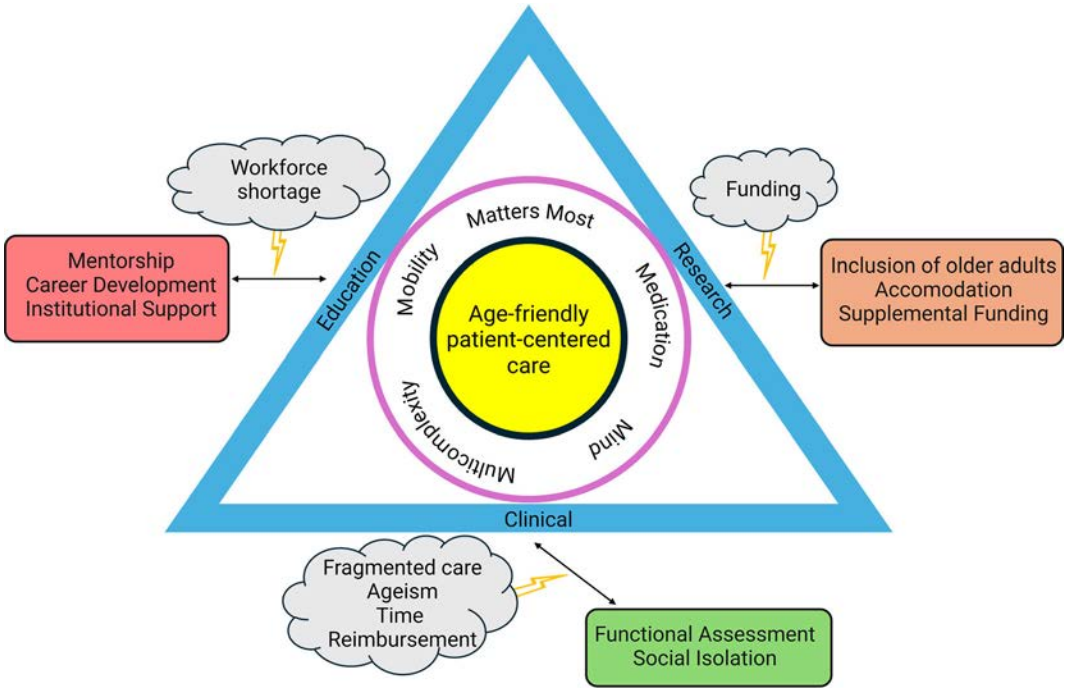


Figure 1. Summary of American Geriatrics Society 2024 clinical, education, and research highlights that align with the Geriatric 5Ms (mind, mobility, medications, what matters most, and multicomplexity) and promote age-friendly patient-centered care. Also listed are potential barriers (clouds) to implementing these programs into the field of rheumatology.

approach to fall screening would add value to the care of older adults with rheumatic diseases, promoting improved patient reporting of fall behaviors and allowing for important clinical interventions or referrals focused on maintaining and improving mobility.

Life-space mobility. Mobility in all individuals, but especially older adults, has implications beyond physical activity itself, including an impact on social participation. The ability to engage in social activities and events is often reported as a key part of older adults' health goals.[20] Life-space mobility (ie, an individual's multidimensional engagement with their environment over time)[21] is not captured by standard measures of activities of daily living. At AGS 2024, C. Barrett Bowling, MD, MSPH,[22] discussed ongoing research on measurement of life-space mobility captured through the Life-Space Assessment. The Life-Space Assessment measures community mobility and social participation, yielding a summative quantitative measure of how far an individual tends to go, how often they go there, and how much help they need to get there.[23] In community-dwelling older adults, life-space mobility is independently associated with mortality[21] and has been shown to decline in people with chronic conditions (eg, chronic kidney disease) or in the setting of acute events.[24] Life-space mobility is an innovative measure that could prove to be relevant in older adults living with rheumatic diseases, allowing for improved assessment of functional mobility and incorporating a measure of "what matters most" to our patients.

Social isolation. Social isolation and loneliness are common among older adults and are associated with numerous adverse health outcomes, including incident cardiovascular disease, stroke, and mortality.[25] Recognizing the significant impact of social isolation and loneliness, further intensified by the COVID-19 pandemic, the US Surgeon General issued an advisory on the epidemic of loneliness and social isolation in 2023.[26] Although social isolation and loneliness have received limited attention to date in rheumatology, they may be amplified in individuals with rheumatic diseases, because of functional limitations, fatigue, and chronic pain,[27] and are expected to have a clinical impact, especially on physical functioning and medication adherence.[28]

A practical approach to loneliness and social isolation was presented at AGS 2024 by Ashwin Kotwal, MD, MS,[29] beginning with (1) identifying loneliness and social isolation through the use of existing tools (eg, University of California Los Angeles 3-Item Loneliness Scale[30] and Berkman-Syme Social Network Index[31]) and conversation; (2) discussing contributions to loneliness and providing support for emotional processing; (3) formulating individualized solutions through shared decision-making, including clinical (eg, addressing pain) and social (eg, social support) interventions; and (4) directing attention to related medical needs (eg, advanced care planning and psychosocial support).[32] Although these efforts should be situated within a multidisciplinary team and engage loved ones and caregivers, awareness of this approach is needed among rheumatologists and allied rheumatology professionals, who often play a prominent role in the care of adults with rheumatic conditions at risk for social isolation.

Educational highlights: fostering the next generation of clinician scientists

Rheumatology fellowship programs are crucial for expanding the workforce and cultivating clinician scientists to address the needs of the aging population. However, barriers such as lack of dedicated time, funding, and training in analytic skills hinder research among trainees.[33] This issue affects various medical specialties, including geriatrics.[34] As part of a session focused on how junior investigators can engage with education, health systems, and community partners to successfully conduct studies with older adults, Ashna Rajan, MD, discussed the GEri-pall Research and Mentorship program at Brown University, designed to equip fellows with mentorship and resources to pursue scholarly work within the timeframe of their fellowship.[35] In its first three years, 95% of fellows had abstracts selected for poster presentation at a national conference, and 30% of these abstracts resulted in manuscripts. Notably, the opportunity to develop mentorship skills and build collaborative networks benefited both the fellows and the faculty involved.

Mentorship programs similarly could help the rheumatology community overcome barriers to academic careers.[33] The Brown GEri-pall Research and Mentoring program begins in the first month of fellowship, providing dedicated time for identifying a hypothesis-driven original research interest. Fellows also are paired with faculty mentors early in the program for regular meetings to set and meet milestones. Additionally, the program provides tangible resources such as a Data Sandbox, a dedicated analyst, and a Continuing Medical Education allowance. The Data Sandbox is innovative because a program-specific institutional review board enables fellows to access data from the beginning of their fellowship, eliminating what is often a rate-limiting step in research.

In rheumatology, national mentorship matching programs like Creating Adult Rheumatology Mentorship in Academia (CARMA) and ACR/Childhood Arthritis and Rheumatology Research Alliance (CARRA) Mentoring Interest Group (AMIGO) developed by the ACR and CARRA support both trainees and junior faculty. However, access to data and analytic support is a significant barrier to research success during training.[33,36] The vast majority of existing rheumatology-focused datasets that have been leveraged in trainee-led research suffer from limited accessibility. The ACR's Rheumatology Information System for Effectiveness (RISE),[37] for example, does not provide for "hands-on training" in data management or analysis, as all data manipulation and analyses using RISE data are conducted centrally at analytic

centers with aggregate output shared with researchers. Although collectively demonstrating substantial value in prior mentored rheumatology research, other national datasets such as those from the Veterans' Affairs Rheumatoid Arthritis Registry (VARA) registry, FORWARD—The National Databank for Rheumatic Diseases, or CorEvitas (among others) are also limited in terms of access and/or analytic support available to trainees and mentors alike.[38] We encourage the rheumatology community to be creative and proactive (seeking institutional support or creating new opportunities at ACR, Rheumatology Research Foundation (RRF), and other rheumatology-focused funding organizations) to enhance training and research by expanding grant funding with access to and analytic support for rheumatology-focused datasets for trainees and/or early career investigators. Emulating the Brown GEri-pall program by making research resources more accessible will support the next generation of clinician scientists to inform the care of the aging population with rheumatic diseases.

Research highlights: building toward an age-inclusive evidence-based care model

Despite being the most frequent consumers of medications, devices, and health care services, older adults are underrepresented in clinical trials.[39] In rheumatology, one-third of clinical trials focused on rheumatoid arthritis and osteoarthritis included an upper age limit for enrollment without providing clear justification.[40] Consequently, clinical practice guidelines are developed based on studies involving mostly middle-aged adults, which is not representative of the rheumatology patient population as a whole. To address this inclusivity gap, the NIH enacted the Inclusions Across the Lifespan Policy,[41] which requires investigators to include individuals of all ages, unless there is a scientific or ethical reason not to include a specific age group.

At AGS 2024, challenges related to and innovative approaches for "geriatricizing" research to build age-inclusive evidence-based clinical care were discussed by C. Barrett Bowling, MD, MSPH.[22] The 5T framework (Target population, Team, Time, Tools, Tips for accommodation)[42] was highlighted as a framework to adapt research efforts to promote the inclusion of older adults in studies.[43] The 5Ts, outlined in Figure 2, were designed to enhance research inclusivity of people across the lifespan in general and likely would enhance rheumatology-based research specifically.

The 5Ts framework can be tailored to support recruitment, enrollment, and retention of older adults in clinical trials and the inclusion of outcome measures that "matter most" to older patients. We believe that adoption of age-friendly frameworks by rheumatology funders, similar to that promoted by the National Institute on Aging (NIA), will enhance inclusion of older adults and improve the relevance and quality of rheumatic disease research. We suggest that including a plan for recruiting and accommodating adults across the age span should be required in all proposals submitted to rheumatology funders (similar to

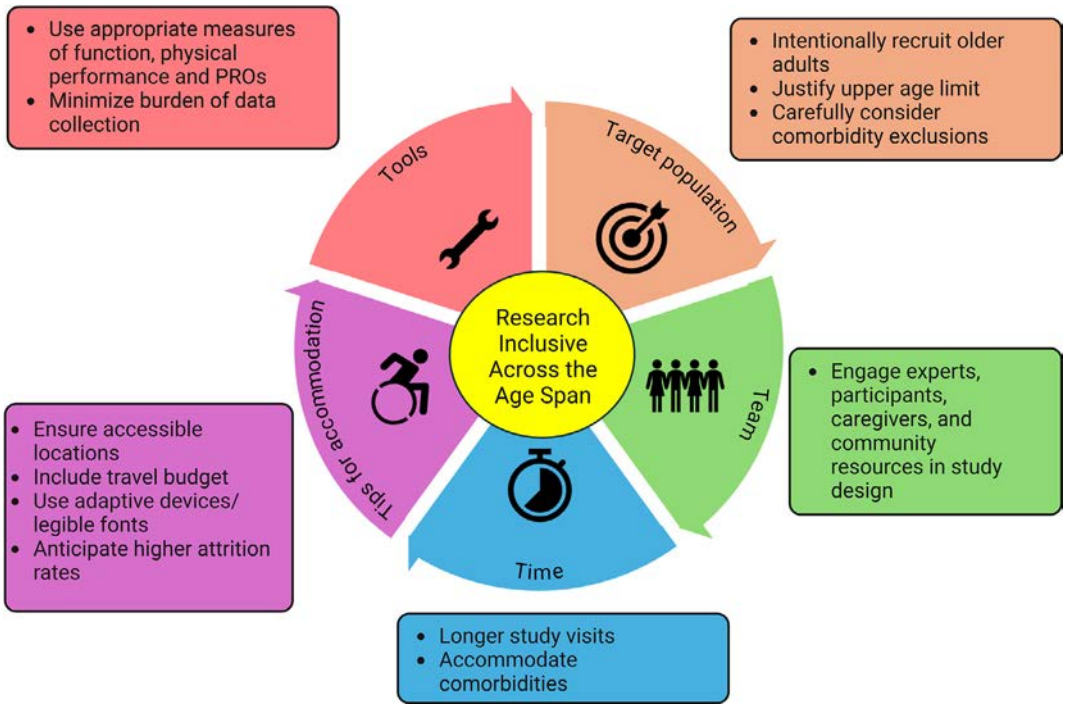


Figure 2. Rationale and points to consider when implementing the 5Ts (target population, team, time, tips to accommodate, tools in rheumatology research) in rheumatology research to ensure representation of older adults. Adapted from Bowling CB, et al. *J Am Geriatr Soc* 2019;67:342–346. PRO, patient-reported outcome. [Correction added on 25 February 2025, after first online publication: Figure 2 has been corrected.] Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25483/abstract.

sex as a biologic variable) and considered when scoring.[44] As it will take time to adopt new grant funding policies, we feel that in the interim funders should consider research supplements for ongoing studies to expand recruitment to include older adults. These efforts are likely to result in improved care delivery to the aging rheumatology patient population.

A case study on implementing geriatric principles to surgical practice: The Geriatric Surgery Verification Program

The American College of Surgeons, AGS, and the John A. Hartford Foundation joined efforts and created the Geriatric Surgery Verification (GSV) program to improve surgical care for older adults by developing a structured program that addresses the goals and needs of older adults. This program targets frail adults 75 years of age or older who are considering in-hospital surgery and aims to optimize their surgical care through prehabilitative interventions based on comprehensive geriatric assessments. The GSV program was developed based on the geriatric 4/5M principles and focuses on improving outcomes in the following four main domains: (1) goals of care and decision-making, (2) cognition screening and delirium, (3) maintenance of function and mobility, and (4) nutrition and hydration optimization. To maximize impact, detailed protocols on patient care, governance, resource allocation, and data surveillance were developed based on expert consensus. These standards enable real-time quality monitoring and improvement in implementation, promote research advancements, and facilitate scaling across surgical specialties and hospital environments.

The success of the GSV program demonstrates the value and potential of multidisciplinary collaboration to develop and disseminate innovative models to improve the care of older adults. Importantly, several health care systems and specialties, such as the Veterans Health Administration, emergency medicine, and cardiology, also have developed age-friendly or age-focused care models.[45] They are primed to meet a new quality measure that CMS will introduce in 2025, based on age-friendly principles, to ensure that hospitals align care with older adults' goals and preferences.[46] How this can be adapted to ambulatory care for older patients with rheumatic conditions has yet to be explored. Further, the rheumatology field is recognizing a shift in quality measurement from assessment of solely disease-specific metrics (eg, focused on rheumatoid arthritis or systemic lupus erythematosus) toward development of cross-cutting, universal quality metrics that reflect what matters most to patients across disease states—also consistent with age-friendly care.[47,48]

Conclusions

Individuals with rheumatic conditions accumulate age-related comorbidities and become frail at a younger age than the general population,[49] heightening the need for age-friendly care.[50] Moreover, older adults with rheumatic diseases have higher health care resource utilization and a greater burden of functional impairment and disability.[51] Hence, it is timely and imperative that we as rheumatologists consider how to adopt geriatric principles and leverage multidisciplinary expertise to develop comprehensive standards of care and move toward value-based care for older adults with rheumatic diseases.

A recent survey conducted by the Hartford Foundation revealed that about 8 in 10 older adults (not all with rheumatic diseases) "feel the health care system is not prepared for the growing and changing needs of our country's aging population."[52] Older adults strongly desired a person-centered holistic approach to care with more frequent discussion of "what matters most to you" that would enable them to stay independent and age in place. Those who received age-friendly care reported better health care relationships and outcomes. Further work is needed in clinical and research settings to implement this care model in rheumatology.

In the last decade, an interest group of rheumatologists dedicated to improving the care of older adults has grown steadily. The group aims to bridge the fields of rheumatology and geriatrics and has collaborated in multiple forums to disseminate geriatric principles to the rheumatology community. To date, the ACR has hosted aging-related sessions on mobility, physical function,[53,54] inflammaging, frailty, multimorbidity, body composition,[55] cognitive impairment, and palliative care[56,57] at ACR Convergence national meetings. Recently, "the intersection between aging and rheumatic disease and the complexities of caring for older adults" was recognized as a topic of interest by the Annual Meeting Planning Committee[58] for ACR 2025 Convergence. The ACR also sponsored an initiative to include measures relevant to aging populations in a special issue of *Arthritis Care & Research* devoted to patient outcomes in rheumatology.[59] The ACR's initiatives and recognition of the importance of improving the care for the aging rheumatology patient population highlights the potential for further collaborative initiatives between the ACR and aging experts. For those interested in pursuing aging research, the AGS and NIH/NIA have educational and funding opportunities that focus on developing investigators at the intersection of specialty care and geriatrics. The community of investigators, leaders, and mentors in aging has welcomed and supported specialists from both the medical and surgical fields.[3]

Caring for older adults with rheumatic disease is deeply rewarding because the approach is guided increasingly by the patient's goals, preferences, and needs.[3] Adopting the patient-centered and collaborative approaches presented at AGS will not only enrich our scientific knowledge, but also foster a holistic understanding of the aging process, ultimately improving patient outcomes and advancing the fields of both rheumatology and geriatrics. As the population with rheumatic diseases ages,

delivery of evidence-based, safe, effective, and patient-centered care should be the standard of care, and ideally delivered in collaboration with geriatricians and other health professionals. Comprehensive training in aging principles for rheumatologists is now more relevant and timelier than ever to enhance the care of older adults with rheumatic diseases.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Levinson J, Buehring B. Pitfalls and pearls in diagnosing inflammatory arthritis in older patients. Joint Bone Spine 2024;91(6):105719.
2. van Onna M, Boonen A. Challenges in the management of older patients with inflammatory rheumatic diseases. Nat Rev Rheumatol 2022;18:326–334.
3. Hurria A, High KP, Mody L, et al. Aging, the medical subspecialties, and career development: where we were, where we are going. J Am Geriatr Soc 2017;65:680–687.
4. Hwang U, Carpenter C, Dresden S, et al; GEAR* and GEAR 2.0** Networks. The Geriatric Emergency Care Applied Research (GEAR) network approach: a protocol to advance stakeholder consensus and research priorities in geriatrics and dementia care in the emergency department. BMJ Open 2022;12(4):e060974.
5. Dale W, Klepin HD, Williams GR, et al. Practical assessment and management of vulnerabilities in older patients receiving systemic cancer therapy: ASCO guideline update. J Clin Oncol 2023;41(26):4293–4312.
6. Rosko AE, Olin RL, Artz A, et al. A call to action in hematologic disorders: a report from the ASH scientific workshop on hematology and aging. J Geriatr Oncol 2018;9:287–290.
7. Forouzandeh F, Alexander K, Forman D, et al. Cardiovascular disease in the older adult: where are we 4 decades later? JACC Adv 2024;3:100820.
8. Goyal P, Kwak MJ, Al Malouf C, et al. Geriatric cardiology: coming of age. JACC Adv 2022;1:100070.
9. Kumar M, Orkaby A, Tighe C, et al. Life's essential 8: optimizing health in older adults. JACC Adv 2023;2:100560.
10. Schwartz JB, Schmader KE, Hanlon JT, et al. Pharmacotherapy in older adults with cardiovascular disease: report from an American College of Cardiology, American Geriatrics Society, and National Institute on Aging Workshop. J Am Geriatr Soc 2019;67:371–380.
11. Magnuson A, Loh KP, Stauffer F, et al. Geriatric assessment for the practicing clinician: the why, what, and how. CA Cancer J Clin 2024;74:496–518.
12. Tinetti M, Huang A, Molnar F. The Geriatrics 5M's: a new way of communicating what we do. J Am Geriatr Soc 2017;65:2115.
13. Mate K, Fulmer T, Pelton L, et al. Evidence for the 4Ms: interactions and outcomes across the care continuum. J Aging Health 2021;33:469–481.
14. Guirguis-Blake JM, Michael YL, Perdue LA, et al. Interventions to prevent falls in older adults: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA 2018;319:1705–1716.
15. Stanmore EK, Oldham J, Skelton DA, et al. Risk factors for falls in adults with rheumatoid arthritis: a prospective study. Arthritis Care Res (Hoboken) 2013;65:1251–1258.
16. Vincenzo JL, Bergen G, Casey CM, Eckstrom E. Reframing Fall Prevention and Risk Management as a Chronic Condition Through the Lens of the Expanded Chronic Care Model: Will Integrating Clinical Care and Public Health Improve Outcomes? Gerontologist 2024;64(6).
17. Hoffman GJ, Ha J, Alexander NB, et al. Underreporting of fall injuries of older adults: implications for wellness visit fall risk screening. J Am Geriatr Soc 2018;66:1195–1200.
18. Eckstrom E, Flaherty E, Johnson TM. Safe mobility: the role of falls assessment and prevention in clinical practice. Abstract presented at: American Geriatrics Society Annual Scientific Meeting; May 2024; Virtual.
19. CDC. STEADI, Older Adult Fall Prevention. Accessed November 6, 2024. https://www.cdc.gov/steadi/index.html
20. Saajanaho M, Rantakokko M, Portegijs E, et al. Personal goals and changes in life-space mobility among older people. Prev Med 2015;81:163–167.
21. Johnson J, Rodriguez MA, Al Snih S. Life-space mobility in the elderly: current perspectives. Clin Interv Aging 2020;15:1665–1674.
22. Bowling, CB. Thomas and Catherine Yoshikawa Outstanding Scientific Achievement for Clinical Investigation Award Lecture: "Geriatricizing" Chronic Disease Research: A Geriatrician's Journey in Collaborative Research. American Geriatrics Society Annual Scientific Meeting; May 2024; New York, New York.
23. Baker PS, Bodner EV, Allman RM. Measuring life-space mobility in community-dwelling older adults. J Am Geriatr Soc 2003;51:1610–1614.
24. Bowling CB, Muntner P, Sawyer P, et al. Community mobility among older adults with reduced kidney function: a study of life-space. Am J Kidney Dis 2014;63:429–436.
25. Donovan NJ, Blazer D. Social isolation and loneliness in older adults: review and commentary of a national academies report. Am J Geriatr Psychiatry 2020;28:1233–1244.
26. Our Epidemic of Loneliness and Isolation: The U.S. Surgeon General's Advisory on the Healing Effects of Social Connection and Community. Washington (DC); 2023. Accessed November 6, 2024. https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf
27. Bay LT, Ellingsen T, Giraldi A, et al. "To be lonely in your own loneliness": the interplay between self-perceived loneliness and rheumatoid arthritis in everyday life: a qualitative study. Musculoskeletal Care 2020;18:450–458.
28. Lu J, Zhang N, Mao D, et al. How social isolation and loneliness effect medication adherence among elderly with chronic diseases: an integrated theory and validated cross-sectional study. Arch Gerontol Geriatr 2020;90:104154.
29. Kotwal, A. Loneliness and social isolation - a growing clinical and public health concern. American Geriatrics Society Annual Scientific Meeting; May 2024; Virtual.
30. Hughes ME, Waite LJ, Hawkley LC, et al. A short scale for measuring loneliness in large surveys: results from two population-based studies. Res Aging 2004;26:655–672.
31. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979;109:186–204.
32. Kotwal AA, Meier DE. A paradigm shift-loneliness as a root cause of symptom distress among older adults. J Am Geriatr Soc 2022;70:2201–2204.
33. Ogdie A, Sparks JA, Angeles-Han ST, et al; American College of Rheumatology Early Career Investigator Subcommittee of the Committee on Research. Barriers and facilitators of mentoring for trainees and early career investigators in rheumatology research: current state, identification of needs, and road map to an inter-institutional adult

rheumatology mentoring program. Arthritis Care Res (Hoboken) 2018;70:445–453.
34. Milani SA, Ahmed A, Rajagopal S, et al. The landscape of geriatric fellow scholarly activity participation: findings from a national survey of program directors. Cureus 2023;15:e47989.
35. Rajan A. Amplifying engagement and impact of your research through partnerships with education, health systems and community organizations: de-alienation of research during fellowship. American Geriatrics Society Annual Scientific Meeting; May 2024; Virtual.
36. Irwin AN, Olson KL, Joline BR, et al. Challenges to publishing pharmacy resident research projects from the perspectives of residency program directors and residents. Pharm Pract (Granada) 2013;11:166–172.
37. 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;68:1866–1873.
38. Mikuls TR, Reimold A, Kerr GS, Cannon GW. Insights and implications of the VA Rheumatoid Arthritis Registry. Fed Pract 2015;32:24–29.
39. Bowling CB, Thomas J, Gierisch JM, et al. research inclusion across the lifespan: a good start, but there is more work to be done. J Gen Intern Med 2023;38:1966–1969.
40. Palmowski A, Buttgereit T, Palmowski Y, et al. Applicability of trials in rheumatoid arthritis and osteoarthritis: a systematic review and meta-analysis of trial populations showing adequate proportion of women, but underrepresentation of elderly people. Semin Arthritis Rheum 2019;48:983–989.
41. NIH policy and guidelines on the inclusion of individuals across the lifespan as participants in research involving human subjects. Accessed November 6, 2024. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-116.html
42. Duke. 5Ts framework. Accessed November 6, 2024. 5tsframework.duke.edu
43. Bowling CB, Whitson HE, Johnson TM II. The 5Ts: preliminary development of a framework to support inclusion of older adults in research. J Am Geriatr Soc 2019;67:342–346.
44. NIH. NIH policy on sex as a biological variable. Accessed November 6, 2024. https://orwh.od.nih.gov/sex-gender/orwh-mission-area-sex-gender-in-research/nih-policy-on-sex-as-biological-variable
45. Institute for Healthcare Improvement. Age-friendly health systems. Accessed November 6, 2024. https://www.ihi.org/networks/initiatives/age-friendly-health-systems
46. Wahl E, Makris UE, Suter LG. Taxonomy of quality of care indicators: tracing the path from clinical practice guidelines to quality measurement and beyond. Rheum Dis Clin North Am 2022;48:601–615.
47. Tinetti ME, McAvay G, Chang SS, et al. Effect of chronic disease-related symptoms and impairments on universal health outcomes in older adults. J Am Geriatr Soc 2011;59:1618–1627.
48. Jacobs DB, Schreiber M, Seshamani M, et al. Aligning quality measures across CMS - The Universal Foundation. N Engl J Med 2023;388:776–779.
49. Lieber SB, Wysham KD, Sattui SE, et al. Frailty and rheumatic diseases: evidence to date and lessons learned. Lancet Rheumatol 2024;6:e881–e891.
50. Lieber SB, Young SR, Shea Y, et al. The lived experiences of older adults with systemic lupus erythematosus: patient perspectives. J Rheumatol 2024:jrheum.2024-0478.
51. Sokka T, Krishnan E, Häkkinen A, et al. Functional disability in rheumatoid arthritis patients compared with a community population in Finland. Arthritis Rheum 2003;48:59–63.
52. Meeting the growing demand for age-friendly care: health care at the Crossroads. Accessed November 6, 2024. https://www.johnahartford.org/images/uploads/resources/The_Growing_Demand_for_Age-Friendly_Care_Report_FINAL.pdf
53. Callahan KE, Boustani M, Ferrante L, et al. Embedding and sustaining a focus on function in specialty research and care. J Am Geriatr Soc 2021;69:225–233.
54. Cummings SR, Studenski S, Ferrucci L. A diagnosis of dismobility--giving mobility clinical visibility: a Mobility Working Group recommendation. JAMA 2014;311:2061–2062.
55. Giles JT, Ling SM, Ferrucci L, et al. Abnormal body composition phenotypes in older rheumatoid arthritis patients: association with disease characteristics and pharmacotherapies. Arthritis Rheum 2008;59:807–815.
56. Makris UE, Buehring B, Yung RL. Tackling multicomplexity in aging patients with RA. The Rheumatologist; 2021.
57. Lee JM, Misra D. Optimizing DMARD use in older adults with rheumatoid arthritis. The Rheumatologist; 2024.
58. Topics of interest by the American College of Rheumatology Annual Meeting Planning Committee. Accessed November 6, 2024. https://assets.contentstack.io/v3/assets/bltee37abb6b278ab2c/blte1db263db09020ec/annual-meeting-ampc-topics-of-interest.pdf
59. Nashi R, Misra D. Special considerations in geriatric populations. Arthritis Care Res (Hoboken) 2020;72 (Suppl 10):731–737.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 564–572
DOI 10.1002/acr.25490

CLINICOPATHOLOGIC CONFERENCE

A 26-Year-Old Man With Systemic Lupus Erythematosus, Disseminated Tuberculosis, and Progressive Right Hemiparesis

Prithivi Raaj Prakash, Ankush Garg, Adil Rashid Khan, Ekamjot Singh, Ajay Garg, Mehar Chand Sharma, Neeraj Nischal, Arvind Kumar, Naveet Wig, and Siddharth Jain

CASE PRESENTATION

History of the present illness

A 26-year-old man, a resident of Northern India, presented with complaints of insidious onset, gradually progressive weakness associated with stiffness of his right upper and lower limbs for the past 2 months. He was evaluated on outpatient basis where a magnetic resonance imaging (MRI) of the brain was done (3 weeks before admission in our hospital) which was reported as having subacute infarcts (Figure 1). In the meantime, while he was being evaluated, he developed new-onset cognitive dysfunction in the form of decreased verbal output and loss of episodic memory since the last one week. There were no sensory symptoms or any history suggestive of cranial nerve, cerebellar, bowel, bladder, or meningeal involvement. In view of worsening neurologic deficits, he was admitted for further evaluation.

Past medical history

The patient had a significant background medical history. He was diagnosed with systemic lupus erythematosus (SLE) with class IV lupus nephritis one year back for which he had been treated with pulse followed by tapering oral glucocorticoids and cyclophosphamide induction therapy (Eurolupus protocol) followed by maintenance azathioprine (150 mg/day; ie, 2.5 mg/kg/day). He had a disease flare in the form of oral ulcers, leucopenia, and worsening renal function despite being compliant to therapy, for which steroid dose was increased (to 0.6 mg/kg/day) and azathioprine was switched to mycophenolate mofetil (MMF). One month after starting MMF (three months before current admission), he was admitted for evaluation of pyrexia of unknown origin, the evaluation of which revealed features of disseminated tuberculosis (TB) involving the lungs (microbiologically confirmed through bronchoalveolar lavage Xpert Mycobacterium tuberculosis/rifampicin positivity), lymph nodes, liver, and spleen (as evidenced by the presence of necrotic mediastinal, hilar, and mesenteric lymph nodes and hypoenhancing lesions in the spleen and liver in the contrast-enhanced computerized tomography scan of the chest and abdomen), for which antituberculous therapy (ATT; isoniazid, rifampicin, pyrazinamide, ethambutol) was initiated, MMF was withheld, and steroid dose was decreased. He developed drug-induced liver injury as a result of ATT, requiring modification of ATT. Gradually, his fever resolved over the next month. His medications at the time of admission included oral prednisolone 15 mg/day, hydroxychloroquine 300 mg/day, ramipril 2.5 mg/day, isoniazid 300 mg/day, rifampicin 600 mg/day, levofloxacin 750 mg/day, and ethambutol 800 mg/day, along with pyridoxine and calcium supplements.

Social and family history

There was no history of autoimmune or neuropsychiatric disease in his family. He was a vegetarian by diet; did not smoke or consume alcohol; had normal sleep, bowel, and bladder habits; and belonged to lower-middle socioeconomic status.

Physical examination

On examination, the patient was conscious but not oriented to time, place, or person. He was afebrile and hemodynamically stable. Pallor, icterus, pedal edema, or lymphadenopathy was

Prithivi Raaj Prakash, MD, Ankush Garg, MD, Adil Rashid Khan, MD, Ekamjot Singh, MD, Ajay Garg, MD, Mehar Chand Sharma, MD, FRC PATH, Neeraj Nischal, MD, Arvind Kumar, MD, Naveet Wig, MD, Siddharth Jain, MD, DM, MRCP: All India Institute of Medical Sciences, New Delhi, India.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25490).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25490.

Address correspondence via email to Siddharth Jain, MD, DM, MRCP, at aiims.siddharth@gmail.com or Siddharth_jain@aiims.edu.

Submitted for publication November 26, 2024; accepted in revised form November 26, 2024.

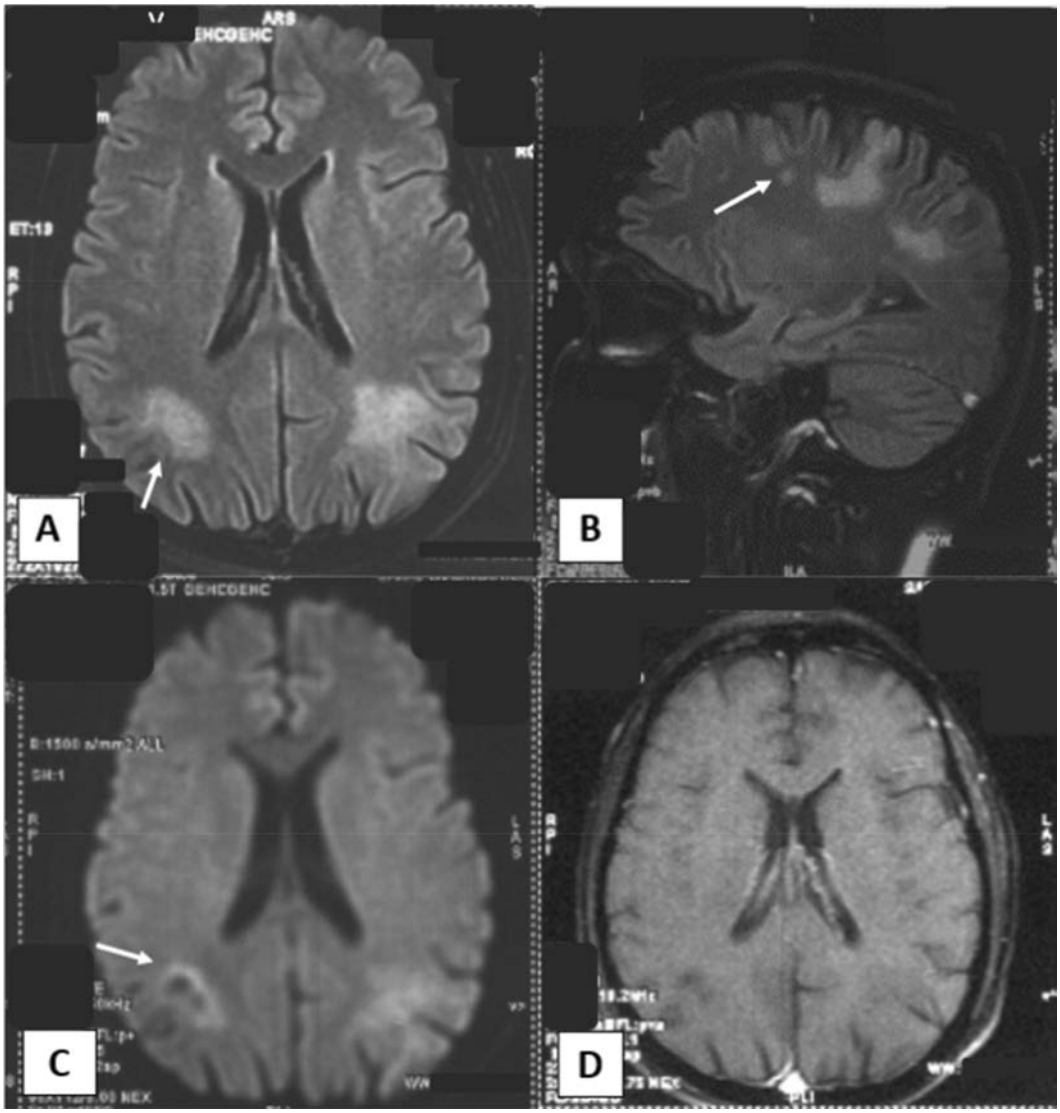


Figure 1. Magnetic resonance image of the brain (three weeks before admission). (A) Axial and (B) sagittal fluid-attenuated inversion recovery (FLAIR) images showing hyperintense focal and confluent lesions involving bilateral parietal, left temporal deep/subcortical white matter without mass effect. These lesions were hypointense on T1 (not shown here); white arrow in (A) shows involvement of subcortical U fibers, and white arrow in (B) shows "milky way sign" (multiple punctate FLAIR hyperintense lesions surrounding the main lesion), which are typical for patients with progressive multifocal leukoencephalopathy. (C) Diffusion-weighted image shows peripheral rim of diffusion restriction (white arrow; "rim and core pattern"). (D) No contrast enhancement of lesions is seen after gadolinium administration. The susceptibility-weighted image did not show blooming, and magnetic resonance angiogram was normal (not shown here).

not present. He had no rash, oral ulcers, joint swelling, or tenderness to suggest lupus disease activity. Although his Glasgow Coma Scale score was 15 of 15, his Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) scores were markedly reduced (13 of 30 and 3 of 30, respectively). On motor examination, he had weakness, spasticity, and hyperreflexia of the right upper and lower limbs, with right ankle clonus and an extensor plantar response. He had a classical pyramidal pattern of weakness, involving the distal group of muscles [Medical Research Council (MRC) grade 2/5] more than the proximal group of muscles (MRC grade 3/5) and affecting extensors more than the flexors in the right upper limb and flexors more than the extensors in the right lower limb. Motor examination of the left upper and lower limbs was normal. Sensory examination, cranial nerves, and cerebellar examination were unremarkable, with absence of meningeal signs. Examination of the abdomen, respiratory, and cardiovascular systems was normal.

Laboratory evaluation

Basic laboratory investigations are summarized in Table 1. Hemogram revealed mild anemia of chronic disease (hemoglobin

Table 1. Laboratory investigations*

Parameter	Value	Normal value
Hemoglobin, g/dL	10.4	12–16
Hematocrit, %	34.8	36–46
Total leukocyte count, n/μL	4,285	4,000–11,000
Differential leukocyte count, N/L, %	58/21	55–70/20–40
Absolute lymphocyte count, n/μL	899	1,000–3,000
Peripheral smear	Mild anisocytosis, predominantly normochromic normocytic cells	–
Iron, μg/dL	87	59–158
Ferritin, ng/mL	520	30–400
TIBC, μg/dL	211	250–450
Lactate dehydrogenase, U/L	162	120–246
Corrected reticulocyte count, %	1.1	0.5–2
Direct Coombs test	3+	–
Urea, mg/dL/	14	17–43
Creatinine, mg/dL	0.5	0.7–1.2
Sodium, mmol/L	140	137–145
Potassium, mmol/L	3.5	3.5–5.1
Total bilirubin, mg/dL	1.1	0.1–1.2
AST, U/L	39	0–35
ALT, U/L	45	0–35
Total protein, g/dL	5.6	6.3–8.2
Albumin, g/dL	2.7	3.5–5
HIV I, II (ELISA)	Negative	–
Hepatitis B surface antigen	Negative	–
Anti-HCV antibody	Negative	–
$CD4^+$ T cell count, n/μL	421	530–1,300
$CD8^+$ T cell count, n/μL	418	350–920
B cell count, n/μL	47	110–570
NK cell count, n/μL	79	20–480
CSF-TC, n/μL	0	0–5
Protein, mg/dL	63	15–60
Glucose, mg/dL	41	50–80
GeneXpert	Negative	–
Bacterial culture	Sterile	–
KOH, fungal culture	Negative, no growth	–
India ink	Negative	–
Cryptococcal antigen	Negative	–
EBV PCR	Negative	–
CMV PCR	Negative	–
VZV PCR	Negative	–
VDRL	Negative	–
JCV PCR (ultrasensitive)	Negative (twice)	–
CRP, mg/L	28.8	3–10
ESR, mm/h	50	<20
C3, mg/dL	111	90–180
C4, mg/dL	29	10–40
24-hour urine protein, mg/d	549	<150
Urine for active sediments	No dysmorphic red blood cells/casts	–
Anti-dsDNA, IU/mL	70	0–100
aCL-IgM, aCL-IgG	Negative	–
Lupus anticoagulant	Negative	–
Anti–β2 glycoprotein 1 antibody	Negative	–
Serum anti-MOG antibody	Negative	–
Serum anti-AQP4 antibody	Negative	–
DVT scan (bilateral lower limbs)	No evidence of DVT	–
2D ECHO	Normal study, left ventricular ejection fraction 55%	–

* 2D, two-dimensional; aCL, anti-cardiolipin; ALT, alanine aminotransferase; AQP, aquaporin; AST, aspartate aminotransferase; CMV, cytomegalovirus; CRP, C-reactive protein; CSF, cerebrospinal fluid; dsDNA, double-stranded DNA; DVT, deep vein thrombosis; EBV, Epstein–Barr virus; ECHO, echocardiography; ELISA, enzyme-linked immunosorbent assay; ESR, erythrocyte sedimentation rate; HCV, hepatitis C virus; JCV, John Cunningham virus; KOH, potassium hydroxide mount; MOG, myelin oligodendrocyte glycoprotein; NK, natural killer; PCR, polymerase chain reaction; N/L, neutrophil/lymphocyte; TC, total count; TIBC, total iron binding capacity; VDRL, Venereal Disease Research Laboratory; VZV, varicella zoster virus.

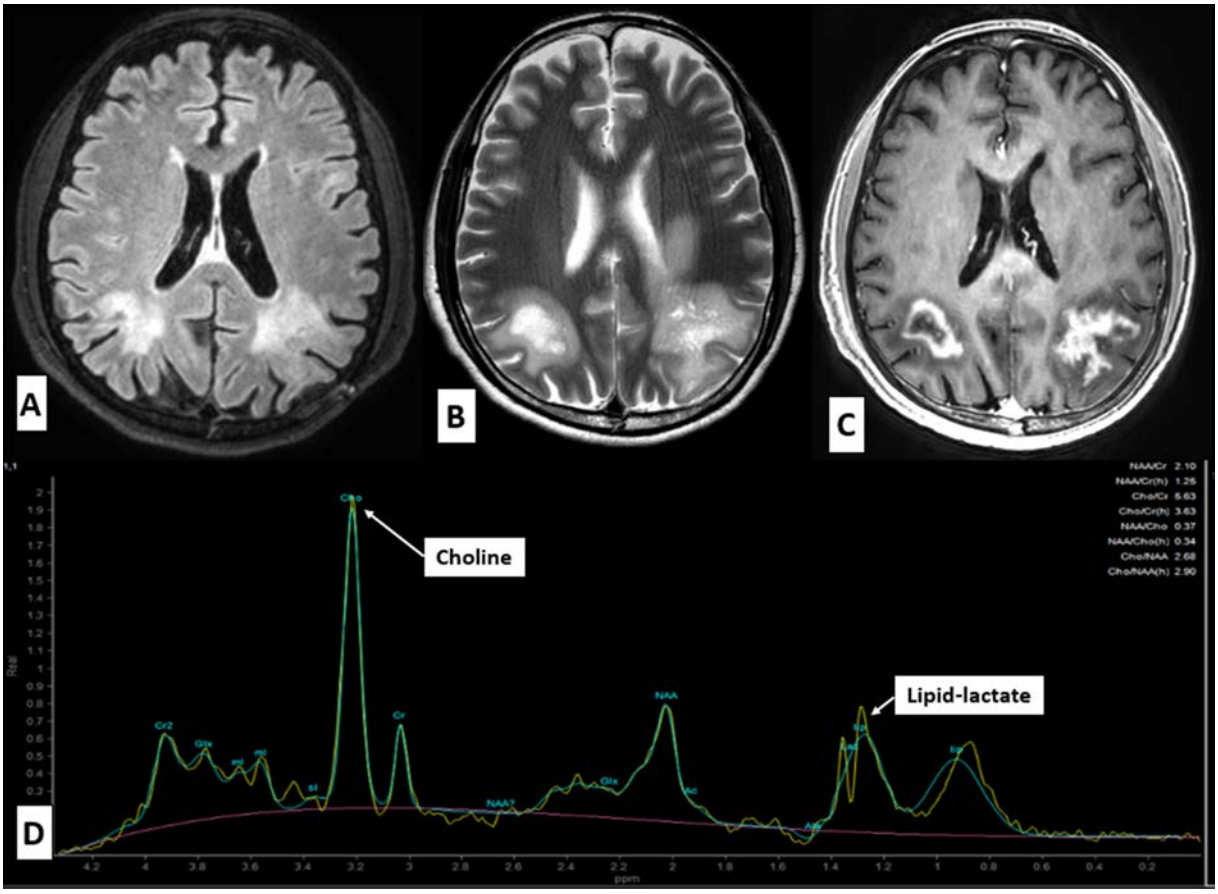


Figure 2. Magnetic resonance image of the brain (at admission): (A) axial fluid-attenuated inversion recovery and (B) axial T2 sequences show hyperintense lesions with increase in lesion size and perilesional edema (indicating active inflammation) with disproportionally less mass effect. (C) Irregular peripheral contrast enhancement after gadolinium administration is seen (not present in Figure 1—suggestive of inflammation). (D) Magnetic resonance spectroscopy shows increased choline (Cho) peak (suggesting active membrane turnover in this patient due to inflammatory activity attributed to progressive multifocal leukoencephalopathy-immune reconstitution inflammatory syndrome) and modestly elevated lipid (lip)/lactate (Lac) peaks. The susceptibility-weighted image showed no blooming, and the diffusion-weighted image showed peripheral rim of diffusion restriction (not shown here). Cr, creatine; Glx, glutamate–glutamine; ml, myo-inositol; sl, scylloinositol.

10.4 g/dL) and mild lymphopenia (899/μL). Renal and liver function tests were normal. Workup for lupus disease activity showed mild proteinuria (24-hour urine protein 549 mg/day) without any active urinary sediments; complement (C3, C4) levels and anti–double-stranded DNA antibody titers were normal, and there was no evidence of ongoing hemolysis, although the direct Coombs's test was positive (Table 1). MRI of the brain done during outpatient evaluation (three weeks before hospital admission) was reviewed and was found to have patchy, ill-defined areas of T2 fluid-attenuated inversion recovery (FLAIR) hyperintense lesions in the right parietal and left frontal, temporal, and parietal lobes with only a peripheral rim of diffusion restriction, no contrast enhancement/edema or any abnormalities in the magnetic resonance angiogram (MRA), making subacute infarcts unlikely (Figure 1). In view of recent neurologic deterioration in the form of cognitive dysfunction, MRI of the brain was repeated at admission (three weeks after the previous MRI), which revealed large T2 FLAIR hyperintense lesions (increased in size compared to previous MRI) in the right parietal and left frontoparietal and left parietotemporal subcortical and deep white matter regions with peripheral diffusion restriction. There was new-onset, irregular contrast (gadolinium) enhancement in these lesions with moderate perilesional edema and no mass effect. Magnetic resonance spectroscopy (MRS) showed elevated choline peak and modestly elevated lipid–lactate peak (Figure 2). MRI of the spine was normal. Cerebrospinal fluid (CSF) analysis showed nil cells, mildly elevated protein (63 mg/dL), and normal glucose levels. A detailed CSF infective workup for bacterial, mycobacterial, fungal, parasitic, and viral infections turned out to be negative (Table 1). CSF John Cunningham virus (JCV) polymerase chain reaction (PCR) by ultrasensitive high-volume extraction technique came out to be negative twice. Serum anti–aquaporin-4 antibody and anti–myelin oligodendrocyte glycoprotein (MOG) were also negative.

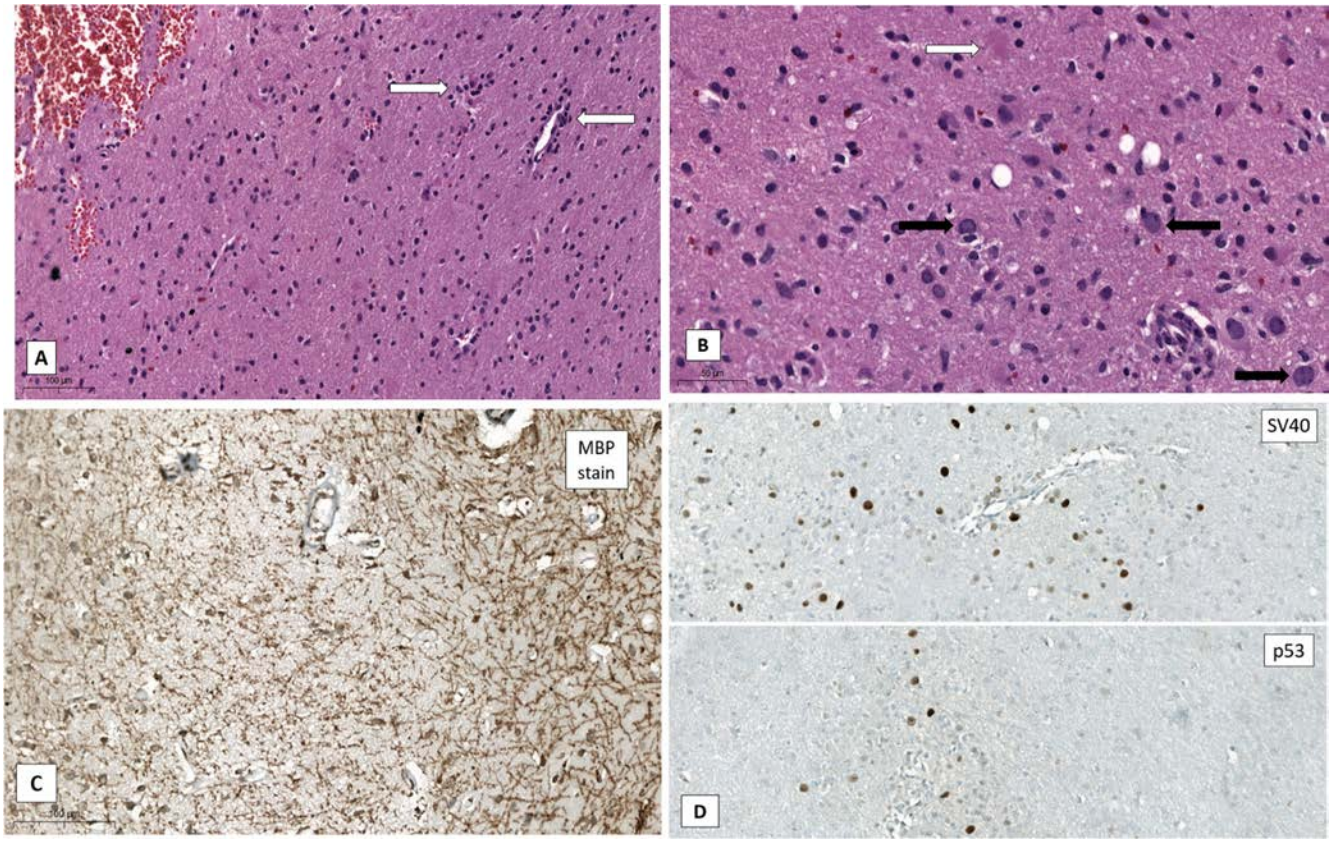


Figure 3. Brain biopsy. (A) Brain parenchyma with focal lymphoid aggregates (white arrows; 200× magnification). (B) Reactive bizarre astrocyte (white arrow) and oligodendrocytes with large intranuclear inclusion and neuronophagia (black arrows; 400× magnification). (C) Myelin basic protein (MBP) immunostaining showing extensive multifocal demyelination (200× magnification). (D) Immunostaining for SV40 and p53 (surrogate markers for John Cunningham virus) showing block positivity in oligodendrocytes (200× magnification). Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25490/abstract.

DIFFERENTIAL DIAGNOSIS

Right-sided hemiparesis with cognitive dysfunction enabled a neuroanatomic localization of the lesion to either the cerebral cortex or the subcortex with adjacent cortical involvement. Taking into account the background of SLE, high-dose immunosuppressive therapy (IST) interrupted by a recent diagnosis of disseminated TB (on ATT), and neuroimaging findings, our initial differential diagnoses were relatively broad and included central nervous system (CNS) TB; other CNS opportunistic infections like nocardiosis, toxoplasmosis, and fungal infections; CNS vasculitis (lupus related); microvascular antiphospholipid syndrome (APS); demyelinating disorders; and CNS malignancy.

CNS infections. In a patient heavily immunosuppressed for underlying lupus disease activity, emergence of new focal neurologic deficits should strongly raise a possibility of a CNS infection. In a high TB-endemic country like India, and a concurrent recent diagnosis of microbiologically confirmed pulmonary TB, dissemination to the CNS (CNS TB) should always be ruled out. However, with the absence of meningeal enhancement, infarcts, abscesses, or tuberculomas in MRI,[1] negative CSF culture and PCR, and onset of symptoms one month after intake of adequate doses of first-line ATT with clinicoradiologic evidence of TB resolution at previously involved sites, the possibility of CNS TB seemed unlikely. Viral, fungal, and parasitic infections of the CNS were ruled out by appropriate serological tests, PCR, cultures, and imaging findings of absence of ring enhancing lesions and meningeal involvement. Progressive multifocal leukoencephalopathy (PML) due to JCV infection was high up in our list of differential diagnoses because the patient had consistent clinical and radiologic features of PML; however, the repeated negativity of ultrasensitive CSF JCV PCR prevented us from reaching a diagnosis of PML.

Autoimmune and vascular etiologies. Neurologic manifestations of SLE, termed as neuropsychiatric SLE (NPSLE), occur in 20% to 40% of patients with lupus[2] and include strokes (which could be due to atherosclerosis, vasculitis, and/or APS), seizures, cognitive impairment, optic neuritis, myelitis, neuropsychiatric manifestations, and peripheral neuropathy.[3] Low disease activity of lupus, negative anti-phospholipid antibodies, normal MRA, and absence of vasculitic infarcts in MRI lowered the

possibility of CNS lupus, which by itself is a diagnosis of exclusion. Subacute nature of illness, absence of hypertensive crisis, and lack of occipital lobe involvement in MRI ruled out posterior reversible leukoencephalopathy syndrome. Neurosarcoidosis most commonly presents with cranial and peripheral neuropathy; focal deficits and neuropsychiatric manifestations are quite rare.[4]

Demyelinating disorders. Absence of optic neuritis, spinal cord involvement, absence of prodromal viral illness/vaccination, and negative anti-MOG and anti–aquaporin-4 antibodies lowered the possibility of acute disseminated encephalomyelitis, neuromyelitis optica spectrum of disorders, and MOG-associated disease. Possibility of tumefactive demyelination (multiple sclerosis) was considered after MRI of the brain due to the presence of large white matter lesions (>2 cm), irregular contrast enhancement, moderate edema, and high choline peak in MRS.[5] However, the sequential changes in the MRI before and after admission were not consistent with tumefactive demyelination.

CNS neoplasia. Primary CNS lymphoma (PCNSL) is seen with increased frequency in autoimmune conditions like SLE, likely related to taking IST.[6] MRI brain features of PCNSL include T2 FLAIR hyperintense, T1 hypo-isointense lesions with homogenous contrast enhancement, edema with/without mass effect, and very high choline peak and high lipid peak in MRS.[7] Although our patient had irregular contrast enhancement and absence of mass effect, PCNSL could not be ruled out without brain biopsy. Metastases to the brain were ruled out by absence of a suggestive history and MRI findings.

Histopathology. Because of lack of conclusive diagnosis after extensive laboratory investigations, detailed CSF workup, and advanced neuroimaging, we proceeded with a brain biopsy from the left parietal lesion, which demonstrated focal aggregates of lymphoid cells, neuronophagia, oligodendroglial intranuclear inclusions, and a few bizarre astrocytes suggestive of PML with evidence of inflammation in the form of lymphoid aggregates (Figure 3). Immunohistochemistry showed positive staining for SV40, p53, and cyclin D1, and JCV PCR of the brain biopsy turned out positive.

THE PATIENT'S COURSE

As per the diagnostic criteria for PML proposed by the American Association of Neurology,[28] our patient was diagnosed with

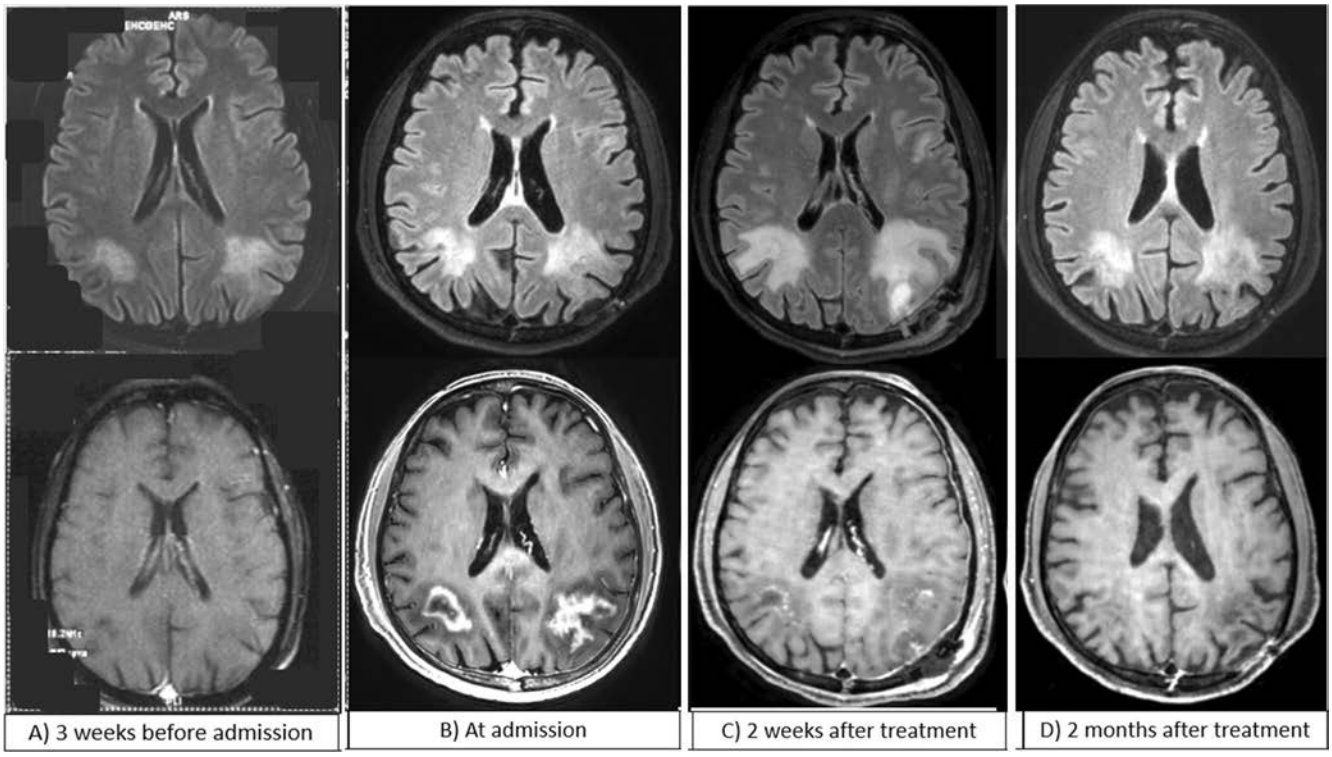


Figure 4. Interval changes in serial magnetic resonance imaging (MRI) of the brain—axial fluid-attenuated inversion recovery (FLAIR; top row) and axial T1 postgadolinium (bottom row)—(A) before admission, (B) at admission, (C) two weeks after treatment, and (D) two months after treatment. Compared to MRI of the brain (A) before admission, (B) MRI of the brain at admission shows an increase in FLAIR hyperintense lesions and irregular postgadolinium contrast enhancement (attributed to progressive multifocal leukoencephalopathy- immune reconstitution inflammatory syndrome). (C) MRI of the brain done two weeks after treatment shows persistence of edema and decrease in contrast enhancement. (D) MRI of the brain two months after treatment initiation shows significant reduction in lesion size, perilesional edema, and contrast enhancement compared to MRI at the time of admission.

definite PML. The underlying disease (SLE) and history of potent IST use (cyclophosphamide, MMF, azathioprine) contributed to an immunocompromised state predisposing to PML. The recent clinical deterioration (in the form of new-onset cognitive decline) accompanied by radiologic worsening in the form of emergence of moderate perilesional edema and contrast enhancement in the recent MRI (Figure 2) was attributed to PML-IRIS, as confirmed by the presence of focal lymphoid aggregates in biopsy (Figure 3).

The patient was started on intravenous dexamethasone 16 mg/day. MRI of the brain was repeated after two weeks, which showed a decrease in contrast enhancement with persistence of edema (Figure 4). He was discharged on a tapering dose of dexamethasone, and subsequently, on follow-up after two months, he had significant improvement in right hemiparesis (modified Rankin scale 2) and cognitive function (MMSE 20 of 30 and MoCA 15 of 30), reaffirming that the clinical deterioration had been due to PML-IRIS. Repeat MRI of the brain after two months showed significant decrease in size, edema, and contrast enhancement of the lesions (Figure 4). At last clinical follow-up, the patient was doing well; dexamethasone had been tapered off and switched to oral prednisolone (7.5 mg/day), and MMF had been restarted. The timeline of significant events before, during, and after admission have been summarized in Supplementary Figure 1.

DISCUSSION

PML is a severe, debilitating, often fatal demyelinating disease of the CNS caused by JCV that is predominantly seen in immunocompromised individuals.[8] Latent infection with JCV from environmental sources occurs in childhood, and around 86% of adults are seropositive for JCV.[9] Impairment of cell-mediated immunity leads to JCV reactivation and rearrangements in the viral genome, making the virus neurotropic, eventually leading to lytic destruction of oligodendrocytes.[8] Historically, HIV infections have contributed to the majority of patients with PML (80%); however, with the advent of effective antiretroviral therapy (ART), the proportion of patients with PML due to HIV has decreased.[10] Hematologic malignancies like chronic lymphocytic leukemia and non-Hodgkin lymphoma contribute to 10% of the PML burden, followed by natalizumab-induced PML in patients with multiple sclerosis comprising 5% of the patients.[8] Other conditions associated with cellular immune dysfunction such as rheumatic diseases, primary immunodeficiency disorders, and immunosuppressant drug use contribute to the remaining minority of patients with PML.[11]

Among patients with rheumatic diseases, two-thirds of PML are reported in SLE[12] attributed mostly to taking IST. In a systematic review of PML in patients with SLE, Henegar et al[13] showed that most patients (66%) had been exposed to some form of immunosuppression before PML onset, but there were no studies that had stratified PML incidence by drug exposure status. There are multiple reports of patients with SLE and PML being treated with biologic agents,[14] but these patients had been concomitantly exposed to other IST as well. The overall incidence of PML in patients with SLE is reported to be 2.4 of 100,000 person-years,[13] with 40% of patients being those who did not take IST, suggesting that inherent lymphopenia or immune dysregulation associated with SLE might contribute to the development of PML.[15] PML in patients with SLE is challenging both from a diagnostic as well as a therapeutic standpoint. Neurologic manifestations of SLE may mimic PML clinically and radiologically, and instances of misdiagnosis have led to catastrophic outcomes because of diametrically opposite management strategies for PML and neuropsychiatric involvement of lupus.[16] This also leads to an underdiagnosis of PML in patients with SLE.[17]

Clinical features like ictal onset and optic nerve/spinal cord involvement are unusual for PML and would favor a diagnosis of NPSLE, whereas MRI evidence of infarcts of different ages (involving both gray and white matter) would favor CNS vasculitis over PML; however, clinicians should be aware that CNS vasculitis and PML can coexist.[13] Multiparametric MRI with MRS provides valuable information regarding tissue composition of various metabolites and helps in differentiation of neoplastic from non-neoplastic intracranial space–occupying lesions. It can preclude the need for brain biopsy in some patients while also helping to improve the yield of targeted brain biopsies.[7] CSF JCV PCR can help in establishing a diagnosis of PML, with a sensitivity of >95% for current assays.[11] Such high sensitivity of test assays virtually obviates the need for an invasive, morbid brain biopsy in most patients. Causes of a false-negative PCR result include intermittent viral shedding in CSF and/or immune recovery or immune reconstitution inflammatory syndrome (IRIS). Although the CSF JCV PCR was twice negative in our patient, a high clinical suspicion and possibility of a false-negative result prompted us to proceed with brain biopsy, which ultimately yielded the diagnosis of JCV PML-IRIS.

No antiviral therapies are available for JCV, and the management of PML rests majorly on immune reconstitution[8] in the form of ART for HIV-PML and withdrawal of IST for non-HIV-PML. Data from patients with HIV-PML have demonstrated a substantial improvement in one-year survival from 10% to 50% with the introduction of ART[18]; however, the data on patients without HIV-PML are very scarce. Apart from withdrawal of IST for PML associated with rheumatic diseases, multiple investigational therapies including mirtazapine, mefloquine, topotecan, cytarabine, cidofovir, and interferon-α have been tried for the management of PML, but none have shown consistent clinical benefit and are thus not recommended.[19] Recently, immune-checkpoint inhibitor (ICI) pembrolizumab use was found to result in clinical improvement and decrease in JCV viral load in five of eight patients with PML (two with HIV-PML and three with non-HIV-PML).[20] Risk of immune-related adverse events has precluded the widespread ICI use in patients with PML associated with rheumatic diseases, although

Lan et al[21] reported a patient with SLE and PML who had a delayed but prolonged response to pembrolizumab.

Although immune reconstitution forms the mainstay of therapy in patients with PML, it is a double-edged sword because it carries a significant risk of triggering IRIS, which further complicates disease management.[8] HIV-PML-IRIS after ART initiation (seen in up to 18% of patients with HIV-PML)[22,23] and PML-IRIS after discontinuation of natalizumab[24] are well-known entities, whereas PML-IRIS in the setting of rheumatic diseases has seldom been reported.[25] A patient with SLE-PML-IRIS occurring in the setting of tofacitinib use was recently reported,[25] and to the best of our knowledge, our patient represents just the second patient with rheumatic PML-IRIS reported in literature thus far. Tackling PML-IRIS requires striking the perfect balance between the need to tamper inflammatory response with glucocorticoids and the risk of PML progression if immunity is fully suppressed.[8] Diagnosis of PML-IRIS requires establishing a clear temporal relationship between immune reconstitution and clinical worsening, detection of JCV in patients with CSF, or brain biopsy and evidence of inflammation in MRI (edema, contrast enhancement) and/or biopsy (lymphoid infiltrates), as proposed by Fournier et al.[22] Although CSF pleocytosis may be expected in IRIS (being an inflammatory condition), there have been prior published reports of PML-IRIS with no CSF pleocytosis in the setting of HIV.[26,27] The published literature, combined with the presence of inflammatory cells on histopathology and treatment response in our patient, suggests that absence of cells in CSF, although rare, does not rule out PML-IRIS. As far as treatment of PML-IRIS is concerned, glucocorticoids have not shown consistent clinical improvement in patients with PML-IRIS and, owing to a potential deleterious role in long-term anti-JCV immune response, are currently recommended only for marked neurologic deterioration with brain swelling.[22] In a review of 54 patients with HIV-PML-IRIS by Tan et al,[23] 12 had been treated with steroids, of which 7 patients showed good neurologic recovery. Patients with favorable clinical response had contrast enhancement on MRI (six of seven patients), earlier steroid initiation (within three weeks of IRIS onset), and longer duration of steroid therapy (mean duration of therapy 13.3 weeks) compared to patients with worse outcomes.[23] Comparison between patients with PML and PML-IRIS has been provided in Supplementary Table 1.

Considering SLE-PML-IRIS, our patient was started on dexamethasone (within one week of IRIS onset), with which he showed significant clinical and radiologic improvement. He was under continuous, close multidisciplinary follow-up to balance the risk of SLE disease flare (due to IST withdrawal) and PML worsening (due to IST reinitiation). His dexamethasone was very slowly tapered and switched to oral prednisolone (7.5 mg/day equivalent), and MMF had been restarted.

This patient's diagnosis and outcome emphasize the need for close vigilance for opportunistic infections in patients with rheumatic diseases on IST; the inadequacies of serological and molecular testing mandating the use of risky, invasive procedures like brain biopsies in patients with difficult diagnoses; and finally, the role of concerted management of these challenging patients through a multidisciplinary team of specialists.

FINAL DIAGNOSIS

Progressive multifocal leukoencephalopathy with immune-reconstitution inflammatory syndrome on a background of SLE with lupus nephritis and disseminated tuberculosis.

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 Jain 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. Bernaerts A, Vanhoenacker FM, Parizel PM, et al. Tuberculosis of the central nervous system: overview of neuroradiological findings. Eur Radiol 2003;13:1876–1890.
2. Hanly JG, Urowitz MB, Su L, et al. Prospective analysis of neuropsychiatric events in an international disease inception cohort of patients with systemic lupus erythematosus. Ann Rheum Dis 2010;69:529–535.
3. The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes. Arthritis Rheum 1999;42:599–608.
4. Joseph FG, Scolding NJ. Neurosarcoidosis: a study of 30 new cases. J Neurol Neurosurg Psychiatry 2009;80(3):297–304.
5. Nakayama M, Naganawa S, Ouyang M, et al. A review of clinical and imaging findings in tumefactive demyelination. AJR Am J Roentgenol 2021;217(1):186–197.
6. Mahale P, Herr MM, Engels EA, et al. Autoimmune conditions and primary central nervous system lymphoma risk among older adults. Br J Haematol 2020;188:516–521.
7. Sawlani V, Patel MD, Davies N, et al. Multiparametric MRI: practical approach and pictorial review of a useful tool in the evaluation of brain tumours and tumour-like lesions. Insights Imaging 2020;11(1):84.
8. Cortese I, Reich DS, Nath A. Progressive multifocal leukoencephalopathy and the spectrum of JC virus-related disease. Nat Rev Neurol 2021;17(1):37–51.
9. Weber T, Trebst C, Frye S, et al. Analysis of the systemic and intrathecal humoral immune response in progressive multifocal leukoencephalopathy. J Infect Dis 1997;176:250–254.
10. Anand P, Hotan GC, Vogel A, et al. Progressive multifocal leukoencephalopathy: a 25-year retrospective cohort study. Neurol Neuroimmunol Neuroinflamm 2019;6:e618.
11. Gomathy S, Panigrahi B, Tirlangi PK, et al. Progressive multifocal leukoencephalopathy in a patient with systemic lupus erythematosus and autoimmune hepatitis. Int J Rheum Dis 2022;25(6):705–713.
12. Molloy ES, Calabrese LH. Progressive multifocal leukoencephalopathy: a national estimate of frequency in systemic lupus erythematosus

and other rheumatic diseases. Arthritis Rheum 2009;60(12):3761–3765.

13. Henegar CE, Eudy AM, Kharat V, et al. Progressive multifocal leukoencephalopathy in patients with systemic lupus erythematosus: a systematic literature review. Lupus 2016;25(6):617–626.
14. Damasio J, Brandao M, Marinho A, et al. Progressive multifocal leukoencephalopathy in lupus. Lupus 2011;20:126.
15. Saylor D, Venkatesan A. Progressive multifocal leukoencephalopathy in HIV-uninfected individuals. Curr Infect Dis Rep 2016;18(11):33.
16. Algahtani H, Shirah B, Othman L, et al. Progressive multifocal leukoencephalopathy misdiagnosed as neuropsychiatric systemic lupus erythematosus with a catastrophic outcome. Neurologist 2022; 27(5):271–275.
17. Cheng CF, Su JJ, Chen YF, et al. Progressive multifocal leukoencephalopathy in a 27-year-old lady with systemic lupus erythematosus - rheumatoid arthritis overlap syndrome. J Formos Med Assoc 2019; 118(11):1560–1565.
18. Antinori A, Cingolani A, Lorenzini P, et al. Clinical epidemiology and survival of progressive multifocal leukoencephalopathy in the era of highly active antiretroviral therapy: data from the Italian Registry Investigative Neuro AIDS (IRINA). J Neurovirol 2003;9(Suppl 1):47–53.
19. Kaplan JE, Benson C, Holmes KK, et al. Guidelines for the prevention and treatment of opportunistic infections in HIV-infected adults and adolescents: recommendations from the CDC, the National Institutes of Health, and the HIV Medicine Association of the Infectious Diseases Society of America. MMWR Morb Mortal Wkly Rep 2009;58(RR04): 1–198.
20. Cortese I, Muranski P, Enose-Akahata Y, et al. Pembrolizumab treatment for progressive multifocal leukoencephalopathy. N Engl J Med 2019;380(17):1597–1605.
21. Lan TY, Chen YS, Cheng CF, et al. Progressive multifocal leukoencephalopathy in systemic lupus erythematosus managed with pembrolizumab: a case report with literature review. Lupus 2021;30(11): 1849–1855.
22. Fournier A, Martin-Blondel G, Lechapt-Zalcman E, et al. Immune reconstitution inflammatory syndrome unmasking or worsening AIDS-related progressive multifocal leukoencephalopathy: a literature review. Front Immunol 2017;8:577.
23. Tan K, Roda R, Ostrow L, et al. PML-IRIS in patients with HIV infection: clinical manifestations and treatment with steroids. Neurology 2009;72(17):1458–1464.
24. Gheuens S, Smith DR, Wang X, et al. Simultaneous PML-IRIS after discontinuation of natalizumab in a patient with MS. Neurology 2012; 78(18):1390–1393.
25. Yun J, Osehobo E, Lawson EC, et al. Tofacitinib-induced progressive multifocal leukoencephalopathy with immune reconstitution inflammatory syndrome. Clin Neurol Neurosurg 2022;214:107143.
26. Tan K, Roda R, Ostrow L, et al. PML-IRIS in patients with HIV infection: clinical manifestations and treatment with steroids. Neurology 2009;72(17):1458–1464.
27. Safdar A, Rubocki RJ, Horvath JA, et al. Fatal immune restoration disease in human immunodeficiency virus type 1-infected patients with progressive multifocal leukoencephalopathy: impact of antiretroviral therapy-associated immune reconstitution. Clin Infect Dis 2002; 35(10):1250–1257.
28. Berger JR, Aksamit AJ, Clifford DB, et al. PML diagnostic criteria: consensus statement from the AAN Neuroinfectious Disease Section. Neurology 2013;80(15):1430–1438.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 573–583
DOI 10.1002/acr.25482

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

25 Years of Biologics for the Treatment of Pediatric Rheumatic Disease: Advances in Prognosis and Ongoing Challenges

Michael Shishov,[1] Pamela F. Weiss,[2] Deborah M. Levy,[3] Joyce C. Chang,[4] Sheila T. Angeles-Han,[5] Ekemini A. Ogbu,[6] Kabita Nanda,[7] Tina M. Sherrard,[5] Ellen Goldmuntz,[8] Daniel J. Lovell,[5] Lisa G. Rider,[9] and Hermine I. Brunner,[5] for the Pediatric Rheumatology Collaborative Study Group Advisory Council

There are over 100 rheumatic diseases and approximately 300,000 children with a pediatric rheumatic disease (PRD) in the United States. The most common PRDs are juvenile idiopathic arthritis (JIA), childhood-onset systemic lupus erythematosus (cSLE), and juvenile dermatomyositis (JDM). Effective and safe medications are essential because there are generally no cures for these conditions. Etanercept was the first biologic therapy for the treatment of JIA, approved in 1999. Since then, other biologic disease-modifying antirheumatic drugs (bDMARDs) and targeted synthetic disease-modifying antirheumatic drugs (tsDMARDs) blocking relevant immunologic pathways have been approved for the treatment of JIA, resulting in a marked improvement of disease prognosis. Conversely, there is only one bDMARD that has been approved for cSLE, but none are approved for the treatment of JDM. Lack of approved therapeutic options, with established dosing regimens and known efficacy and safety, remains a central challenge in the treatment of all PRDs, including autoinflammatory diseases, and for complications of PRDs. This review provides an overview of bDMARD and tsDMARD treatments studied for the treatment of various subtypes of JIA, summarizes information from bDMARD studies in other PRDs, with a focus on pivotal trials that led to regulatory approvals, and highlights improved outcomes in patients with JIA with the reception of these newer medications. Further, we outline barriers and challenges in the treatment of other PRDs. Last, we summarize the current regulatory landscape for bDMARD studies and medication approvals for patients with PRDs.

Introduction

There are over 100 rheumatic diseases and approximately 300,000 children with a pediatric rheumatic disease (PRD) in the United States.[1] Although there are no cures for PRDs to date, early diagnosis, followed by timely, effective therapies are essential for superior disease outcomes. Thus, access to medications with proven efficacy and acceptable safety profiles is crucial. The most common systemic immune-mediated PRDs include juvenile idiopathic arthritis (JIA),[2] childhood-onset systemic lupus erythematosus (cSLE),[3] and juvenile dermatomyositis (JDM).[4] Historically, PRDs such as JIA were associated with poor functional outcomes, the need for joint replacement, and use of functional devices to enable affected patients to perform activities of daily living. Especially for JIA, treatment options have improved markedly since the first biologic disease-modifying antirheumatic drug (bDMARD) was approved for the treatment of JIA in 1999. There are now several bDMARDs and targeted synthetic disease-modifying antirheumatic drugs (tsDMARDs) that have been approved for treating PRDs by regulatory agencies. This has led to a profound change in the practice of pediatric rheumatology and improved the prognosis of PRDs. Nonetheless, unmet

Dr Rider's work was supported by the intramural research program of the National Institute of Environmental Health Sciences, NIH (grant ZIA-ES-101081).

[1]Michael Shishov, MD: Phoenix Children's Hospital, Phoenix, Arizona; [2]Pamela F. Weiss, MD: The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [3]Deborah M. Levy, MD: The Hospital for Sick Children and The University of Toronto, Toronto, Canada; [4]Joyce C. Chang, MD: Boston Children's Hospital, Boston, Massachusetts; [5]Sheila T. Angeles-Han, MD, Tina M. Sherrard, BA, Daniel J. Lovell, MD, Hermine I. Brunner, MD: Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [6]Ekemini A. Ogbu, MD: Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, and Johns Hopkins University, Baltimore, Maryland; [7]Kabita Nanda, MD: Seattle Children's Hospital and University of Washington School of Medicine; [8]Ellen Goldmuntz, MD: National Institute of Allergy and Infectious Diseases, NIH, Rockville, Maryland; [9]Lisa G. Rider, MD: National Institute of Environmental Health Sciences, NIH, Bethesda, Maryland.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25482.

Address correspondence via email to Hermine I. Brunner, MD, at hermine.brunner@cchmc.org.

Submitted for publication July 22, 2024; accepted in revised form December 2, 2024.

needs remain for effective and safe therapies for PRDs, especially the treatment of JIA-associated uveitis, macrophage activation syndrome (MAS), and less prevalent PRDs such as cSLE and JDM. Given the rarity of some PRDs, specialized research networks like the Pediatric Rheumatology Collaborative Study Group (PRCSG) are essential to realize the potential of novel drug targets to become treatments that improve both life and function of children with PRDs.[5] In this review, we will describe therapeutic advances since the advent of approved bDMARDs and tsDMARDs and the regulatory landscape of drug development for PRDs. Notably, we do not discuss issues surrounding the availability of biosimilars and refer the reader to the package insert for the exact labeling language and the review of bDMARD and tsDMARD safety.

Tumor necrosis factor inhibitors are currently considered first line bDMARDS for polyarticular course of JIA

The first bDMARD studied in JIA was etanercept, a fusion protein that inhibits tumor necrosis factor (TNF)-α, and to some extent, TNF-β, by blocking its interaction with cell-surface TNF p75 receptors. In fact, etanercept was the first bDMARD used for the treatment of PRDs that achieved regulatory approval by the US Food and Drug Administration (FDA) under a "fast-track" approach. In 1999, etanercept was approved worldwide to treat polyarticular-course JIA (pcJIA) based on the results of a phase II/III placebo-controlled randomized withdrawal trial.[6] A large phase IV JIA registry of etanercept coordinated by the PRCSG documented the long-term efficacy, safety, reception in younger children, and benefits on quality of life.[7] Because active uveitis was an exclusion criterion for participation in the etanercept JIA clinical trial program, no information regarding the effectiveness of etanercept for the treatment of JIA-associated uveitis was collected. Rather, postapproval reports of uveitis development while receiving etanercept treatment, and lack of uveitis improvement with etanercept, later provided sufficient evidence to establish that etanercept is ineffective for the treatment of uveitis.[8]

Other TNF inhibitors (TNFis) approved by the FDA and the European Medicines Agency (EMA) for the treatment of JIA are monoclonal antibodies (MABs). These include adalimumab and intravenous (IV) golimumab, which were approved to treat pcJIA in 2008 and 2020, respectively. Notably, although widely used in clinical practice, IV infliximab and subcutaneous (SC) golimumab have not achieved FDA approval for treatment of JIA after clinical trials in JIA failed to meet their primary endpoints[9,10]; the former likely failed because the evaluated dose of infliximab was too low. The study of certolizumab in pcJIA (NCT01550003) has been completed, but the results are unavailable. All TNFis studied in pcJIA appear to have similar efficacy and safety as in adults with rheumatoid arthritis (RA; Table 1).[11]

Following a randomized, double-blinded, placebo-controlled study of 90 patients aged 2 to <18 years with active JIA-associated anterior uveitis, the FDA expanded the approval of adalimumab in 2018 for the treatment of noninfectious intermediate uveitis, posterior uveitis, and panuveitis to include pediatric patients ≥2 years old. Lack of approval for anterior uveitis is a major gap in JIA-associated anterior uveitis. Study results showed that adalimumab significantly decreases the risk of treatment failure by 75% relative to placebo.[13]

Other bDMARD classes approved for pcJIA

Tocilizumab, a humanized MAB to the soluble interleukin (IL)-6 receptor that binds to both the soluble and membrane-

Table 1. bDMARDs approved for the various categories of JIA by the FDA*

bDMARD targets	bDMARDs approved for children with JIA and adults with inflammatory arthritis in the United States (year of approval in children)	bDMARDs approved only for adults with inflammatory arthritis (year of initial approval)
TNF-alpha	Etanercept (1999); adalimumab (2008); IV golimumab (2020)	Infliximab (1999); certolizumab (2009)[a]; SC golimumab (2009)
IL-1	Canakinumab (2013)	Anakinra (2001)[b]
IL-6	IV tocilizumab (2011); SC tocilizumab (2018); sarilumab (2024)[c]	–
CTLA4	IV abatacept (2008); SC abatacept (2017)	–
CD20	–	Rituximab (2006)
IL-17	SC secukinumab (2021)	IV secukinumab (2023)
IL-12/IL-23	Ustekinumab (2022)	Ixekizumab (2017)[a]; guselkumab (2020)[a]; risankizumab (2022)[a,d]
PDE4	–	Apremilast (2014)
Janus kinases (JAKs)	Tofacitinib (2020); upadacitinib (2024)	Baricitinib (2018)[b]

* bDMARD, biologic disease-modifying antirheumatic drug; CTLA, cytotoxic T lymphocyte–associated protein; FDA, Food and Drug Administration; IL, interleukin; IV, intravenous; JIA, juvenile idiopathic arthritis; PDE, Phosphodiesterase 4 phosphodiesterase; SC, subcutaneous; TNF, tumor necrosis factor.
[a] Studies in JIA are ongoing.
[b] These bDMARDs were approved for treatment in children with JIA in Europe in 2023 but not in the United States.
[c] Sarilumab was only approved for treatment in adult-sized children with polyarticular-course JIA.[12]
[d] Risankizumab blocks only anti-IL-23.

bound IL-6 receptor,[14–16] is available in IV and SC formulations. The IV formulation was approved to treat pcJIA in 2013 after it had achieved approval to treat systemic JIA (sJIA) in 2011; the SC formulation was approved to be prescribed to patients with JIA in 2018 (Table 1). There are several trials in pcJIA[17] in which tocilizumab was demonstrated to be effective and safe even when used in the long term, and it also facilitated catch-up growth and slowing of radiographic progression.[18,19] Approximately half of the patients in the IV pcJIA trial received methotrexate background therapy and seemed to experience somewhat better treatment response.[16] Safety and efficacy of SC and IV tocilizumab were comparable, other than mild to moderate injection site reactions (ISRs) with SC tocilizumab, which occurred in about 30% of patients in the clinical trial.[16] There are several other IL-6 inhibitors, but currently, tocilizumab remains best studied in PRDs. Although a phase II open-label study of sarilumab is ongoing (NCT02991469), sarilumab was approved in June 2024 for a small subset of patients with pcJIA whose weight is 63 kg or higher.

IV abatacept was approved in 2008 for the treatment of pcJIA, and subsequently, SC abatacept was approved for pcJIA in 2017. Abatacept is a fully human, soluble fusion protein comprised of the extracellular domain of cytotoxic T lymphocyte–associated protein 4 and the Fc component of IgG1, which selectively inhibits the costimulatory signal necessary for full T cell activation by binding to CD80 and CD86, thereby blocking their interaction with CD28. Both IV and SC formulations of abatacept have been studied in two phase III trials in children with pcJIA.[20,21] The efficacy of SC and IV abatacept is comparable for treating pcJIA. The same holds true for safety, except for ISRs, which occur in around 7% of participants receiving SC abatacept.[21] The current label states that concomitant reception of TNFis should be avoided. Systematic reviews with meta-analyses suggest that abatacept therapy is associated with few treatment-emergent infections, with similar responses to therapy and reduction of flare risk in patients with pcJIA compared to other bDMARDs.[22,23]

bDMARDs for sJIA. The introduction of IL-6 and IL-1 inhibitors has changed the landscape of sJIA treatment, allowing for decreased reception of glucocorticoids over the long term, resulting in decreased glucocorticoid-related side effects such as growth delay, reduced bone mineral density, and infections.[24,25] The safety and effectiveness of SC and IV tocilizumab for the treatment of sJIA are comparable.[16,25,26] Episodes of MAS have been observed with IL-6 and IL-1 inhibitors.[16,25,26]

The IL-1 inhibitors, including canakinumab, anakinra, and rilonacept, are another class of bDMARDs that are highly effective in treating sJIA. Canakinumab is an MAB blocking IL-1β and was FDA approved for patients with sJIA who were two years of age or older in 2013, based on the results of two double-blind randomized placebo-controlled trials that provided evidence for rapid response of systemic and articular features in patients with sJIA (Table 1).[27,28] ISRs with canakinumab reception were uncommon and generally mild. Anakinra is a human recombinant form of IL-1 receptor antagonist and blocks the biologic activity of IL-1α and IL-1β by competitively inhibiting IL-1 binding to the IL-1 type I receptor, which is expressed in a wide variety of tissues. The first randomized controlled trial of anakinra, which was four weeks in duration, supported its efficacy for the treatment of sJIA.[29] Additionally, dose escalation of anakinra may yield additional therapeutic benefits.[30,31] However, anakinra, in contrast to canakinumab, is not FDA approved for the treatment of sJIA; this was likely due to low enrollment in the clinical trial, which may have been in part due to the FDA's requirement to include a placebo arm in the trial.

Rilonacept, which binds and neutralizes IL-1, was also studied in a double-blind, placebo-controlled study in sJIA by the PRCSG and the Childhood Arthritis and Rheumatology Research Alliance (CARRA), respectively.[32,33] Although not approved to treat sJIA due to not meeting its primary endpoint, rilonacept was approved for cryopyrin-associated periodic syndrome (CAPS) in 2008. A meta-analysis demonstrated that rilonacept was inferior to canakinumab and tocilizumab in the treatment of sJIA, whereas the risks of serious adverse events were similar among the various bDMARDs.[23] As with IL-6 inhibitors, episodes of MAS occurred with canakinumab treatment.[34] Conversely, anakinra has been suggested for the treatment of MAS associated with active sJIA.[35] The role of bDMARDs in contributing to the increased development of interstitial lung disease in patients with sJIA remains under investigation.[36]

bDMARDs for juvenile psoriatic arthritis. Juvenile psoriatic arthritis (JPsA) was removed from the FDA's waiver list for pediatric studies on October 10, 2023.[37] This means studies in children with JPsA are no longer considered highly impractical and are now subject to the usual requirements for pediatric research, as detailed later in this review. Since then, there have been four FDA-approved medications for JPsA treatment: IV golimumab, ustekinumab, secukinumab, and upadacitinib. Notably, although approved for psoriatic arthritis (PsA) in adults and commonly used in clinical practice, the TNFi adalimumab and SC golimumab are not approved for the treatment of JPsA. However, using exposure matching of dosing data in the pediatric population,[38] IV golimumab was FDA approved for JPsA based on a positive study on pcJIA[39,40] and demonstrated effectiveness and safety in patients with PsA.

Ustekinumab is a fully humanized MAB that binds the p40 subunit of IL-12 and IL-23, thereby inhibiting the interaction of these cytokines with the IL-12Rβ1 receptor and inhibiting their signaling. Without conduct of a dedicated study in children with JPsA, ustekinumab received FDA approval to treat JPsA in 2022 based solely on exposure matching of dosing data from children with other diseases and known efficacy in adults with PsA.[38]

Additional clinical trials of ustekinumab for treatment of JPsA are ongoing to support labeling of ustekinumab in other countries outside the United States (NCT05252533 and NCT05083182).

Secukinumab is a high-affinity, fully human IgG1 MAB that selectively binds and neutralizes IL-17A. SC secukinumab was studied for treatment of JPsA and enthesitis-related arthritis (ERA) in a double-blind, placebo-controlled, phase III, randomized withdrawal trial and was shown to be effective and safe.[41] ISRs with secukinumab reception were rare and mild. Based on this study, the FDA approved SC secukinumab for treating JPsA and ERA in December 2021. Finally, a clinical trial of children with JPsA receiving SC risankizumab, an IL-23 inhibitor, is underway (NCT06100744).

bDMARDs for ERA. Drug development and approval for children with JIA falling into the ERA category of JIA has been hampered for more than a decade. Indeed, patients with ERA were often not included in clinical trials that generally focused on children with pcJIA. Despite off-label TNFi prescription at doses approved for patients with pcJIA, and their consideration as first-line treatment of ERA, per current treatment guidelines,[42,43] none of the currently marketed TNFis have FDA approval for treating ERA. Based on the available scientific evidence,[43] the FDA recently considered ERA as the pediatric counterpart of nonradiographic axial spondyloarthropathy (nr-axSpA). Since then, the FDA approved secukinumab for treating ERA in 2021. Like adults with nr-axSpA or ankylosing spondylitis, children with ERA experienced improvement in signs and symptoms of their disease with the reception of SC secukinumab, which was comparable to that of children with JPsA.[41] With secukinumab treatment, there was an improvement of both joint inflammation and associated enthesitis.

tsDMARDs for JIA

In recent years, tsDMARDs, also known as oral small-molecule inhibitors that affect intracellular signaling pathways, have been studied in patients with JIA. Currently, tsDMARDs for JIA are only approved for pcJIA and JPsA. EMA approval for baricitinib took place in 2023, whereas FDA approval occurred for the oral JAK inhibitors tofacitinib in 2020, and for upadacitinib in 2024, based on the results of randomized placebo-controlled withdrawal design studies.[44–46] Although there are significant boxed warnings, including cardiovascular events and thrombosis in adults receiving tsDMARDs, it is not clear that these serious adverse safety warnings apply to children.

bDMARDs for cSLE

Belimumab is an MAB against B lymphocyte stimulator (B lymphocyte activating factor) and is available as an IV and SC preparation. Despite initial approval of belimumab in 2011 for adults with systemic lupus erythematosus (SLE), belimumab was not approved for cSLE until 2019 or for childhood-onset lupus nephritis (LN) until 2022. The latter approval was, again, based solely on extrapolation from adult LN data and the available data from cSLE, also using Bayesian analysis.[47,48]

A double-blind, placebo-controlled, parallel randomized trial enrolled only 93 patients very slowly and hence was not powered to show efficacy. At week 52, there were numerically more Systemic Lupus Responder Index 4 responders in patients receiving IV belimumab compared to those receiving placebo (52.8% vs 43.6%; odds ratio 1.49, 95% confidence interval 0.64–3.46).[47] Based on the results of an open-label study with the primary endpoints of pharmacokinetics and pharmacodynamics in cSLE, SC belimumab was approved to treat cSLE in children aged five years or older in May 2024.[49] In 2021, anifrolumab was approved for the treatment of adults with SLE and/or LN. Since the study in children began in 2023 (NCT05835310), a delay in potential FDA approval of anifrolumab for several years is likely.

bDMARDs in other PRDs: monogenic autoinflammatory diseases

Dramatic responses to anakinra in selected patients with sJIA, CAPS, and deficiency of the IL-1 receptor antagonist (DIRA) provide evidence for pediatric application. Indeed, anakinra was approved by the FDA for the treatment of DIRA in 2020. Two consecutive phase III studies of 47 adult patients with CAPS supported the effectiveness of rilonacept in this group of autoinflammatory diseases.[50] Likewise, rilonacept has been shown to be effective for the treatment of familial Mediterranean fever (FMF), decreasing the rate of flares by 41% (equal-tail 95% credible interval 39%–85%), but regulatory approval is not in place.[51,52] In addition to CAPS, canakinumab was approved in 2016 for the treatment of monogenic autoinflammatory diseases, such as FMF, mevalonate kinase deficiency, and TNF receptor–associated periodic syndrome, using a basket trial design.[53] The most frequently reported adverse event was infections, with a few serious infections.[53]

bDMARDs in other PRDs

Rituximab is an anti-CD20 MAB that depletes pre–B lymphocytes and mature B lymphocytes but not plasma cells or stem cells. Rituximab is used off label to treat many PRDs, including cSLE, but only holds FDA approval for pediatric antineutrophil cytoplasmic antibody–associated vasculitis (AAV) for patients two years of age and older in combination with glucocorticoids. Rituximab was studied in a randomized, placebo-phase clinical trial in adults with dermatomyositis (DM) and polymyositis and children with JDM.[54] Despite the failure to meet the primary endpoint of time to improvement in the myositis response criteria, subsets of patients appeared to respond to rituximab,

including children with JDM and patients with certain myositis autoantibodies.[54]

Mepolizumab is an MAB targeting IL-5 and is FDA approved for adult patients with eosinophilic granulomatosis with polyangiitis (EGPA) but not in children with EGPA. Of note, mepolizumab is FDA approved for patients ages 12 years and older with hypereosinophilic syndrome. Finally, avacopan, a tsDMARD that is a complement 5a receptor antagonist, was FDA approved as an adjunctive treatment in adults with AAV in 2021, but the indication was not extended to pediatric patients despite inclusion of adolescents 12 years or older in the clinical trial program. However, a clinical trial in pediatric patients with AAV is planned (NCT06321601).

None of the currently marketed bDMARDs have been approved for the treatment of other rarer PRDs or nonmonogenic autoinflammatory diseases. Such diseases include JDM and juvenile systemic sclerosis. In 2021, the FDA granted approval to Octapharma USA for Octagam 10% as the first, and only, intravenous immunoglobulin (IVIG) to be indicated for the treatment of adult DM, based on the results of a randomized, blinded clinical trial.[55] IVIG has not been formally studied in children with JDM, and Octagam® 10% is not licensed for the treatment of JDM.

Improvement of disease outcomes with JIA since introduction of bDMARDs

Before the introduction of bDMARDs, children with JIA experienced marked pain and reduction of well-being and physical function from their disease despite anti-inflammatory therapy.[56] Because most bDMARDs have received indications for the treatment of JIA, there are now robust epidemiologic data describing the impact of bDMARDs in children with JIA. Reports of JIA five-year outcomes in the pre-bDMARD era supported that 38% of children continued to have active synovitis despite anti-inflammatory therapy and that 16% of the children experienced moderate to severe impairment of physical functioning, including impaired school functioning and activities of daily living.[57] Disability, as measured by the Childhood Arthritis Health Assessment Questionnaire (CHAQ; range 0–3, 3 being worse physical function), in patients with pcJIA was at a median (range) CHAQ score of 0.31 (0–3), mean (±SD) pain score of 3.1 (0-3.44) as measured on a visual analog scale (0–10, 0 being no pain), and the mean (range) of active joints with arthritis was 8.1 (0–42).[58,59] However, the authors suggest that the full impact of the changed landscape of treatment options for JIA is better evaluated by using large registries such as those assembled by the Pediatric Rheumatology Care and Outcomes Improvement Network (PR-COIN) and CARRA.

Conversely, in contemporary cohort studies in which 25% to 45% of children with JIA are treated with bDMARDs, the mean (±SD) CHAQ scores are 0.2 (±0.4), mean (±SD) pain ratings are 1.9 (±2.5), and the mean (±SD) number of joints with active arthritis is 1.0 (±2.7), indicating over 90% improvement of disability, about 40% improvement in pain, and 90% improvement in joints with active arthritis.[60] Taken together, the approval of bDMARDs undoubtedly led to a profound improvement of JIA disease control, overall prognosis, and a change in treatment paradigms, with PR-COIN further improving the impact of bDMARD access through its treat-to-target and implementation activities.[61] However, it is worth noting that one should use caution in comparing data among pediatric bDMARD studies because the design of the trials and included patient populations differ. Unfortunately, continued pain and disability exist, especially in children with JPsA and ERA, likely because of limited access to approved bDMARDs until recently.[60]

Overall, considering the continuous burden of PRDs, and despite access to advanced medications in the past 25 years, there is an ongoing need for additional medications, as was highlighted by both the PRCSG and CARRA.[62] For example, many children with pcJIA fail to achieve clinical remission, there are only limited therapeutic choices for ERA and cSLE, and there are no approved medications for JDM.[62] Other concerns include limited efficacy and tolerability of bDMARDs, as reflected by the frequent need to switch bDMARDs. In a real-world observational study of patients with JIA who received bDMARD therapy observed for an average of 3.7 years, 24% and 7%, respectively, were prescribed a second and third bDMARD.[63] Furthermore, children enrolled in JIA medication trials whose TNFi or IL-6 inhibitor treatment failed, but not abatacept treatment,[6] had a 15% to 25%[7] lower likelihood of responding to subsequent treatments for reasons not well understood.

Many PRDs are orphan diseases that occur in fewer than 200,000 persons in the United States, for which there is no reasonable expectation that pharmaceutical companies will recover the costs incurred during drug development following pharmaceutical sales. The Orphan Drug Act of 1983 incentivized pharmaceutical companies to develop drugs for rare diseases by offering market exclusivity and tax credits.[64,65] Although anakinra was approved for the treatment of rare diseases in children, such as CAPS and DIRA, the Orphan Drug Act does not include the requirement of pharmaceutical sponsors to conduct studies in children when the adult counterpart is an orphan disease. Examples include the approval of IVIG for adults with DM and avacopan for adults with AAV, but no studies have been conducted in children with these two related PRDs, and these medications are not approved for children with these conditions. The Inflation Reduction Act requires Medicare price negotiation for selected drugs beginning in 2026.[66] This law will exempt orphan drugs from negotiations, although this exemption applies only to orphan drugs with a single approved indication. Thus, drug companies may have reduced incentives to pursue additional uses of their orphan products going forward. Fortunately, there are remaining incentives for companies, such as the Rare Pediatric Disease Designation and Priority Review voucher program.

Regulations relevant for the approval of drugs for children with PRDs

The FDA is the federal agency charged with overseeing drug manufacturing, labeling, and safety of medications and biologic products in the United States. The Food, Drug, and Cosmetic Act (FD&C Act) requires "substantial evidence" resulting from "adequate and well-controlled investigations" to demonstrate that a new drug "will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the proposed labeling." New drug and biologic applications are submitted to, and reviewed by, the FDA for potential approval for marketing in the United States.

In the United States, medication labeling and drug studies are governed mainly by two sections of the FD&C Act: section 505A pertains to the Best Pharmaceuticals for Children Act (BPCA)[67] and section 505B to the Pediatric Research Equity Act (PREA),[68] respectively.[69,70] These two laws encourage and/or require drug companies to study their products in children. The BPCA provides pharmaceutical manufacturers with six months of additional market exclusivity after the completion of pediatric drug studies requested by the FDA. There is no exclusion for orphan diseases, but biologic therapies are not covered by this law. The PREA necessitates new drugs to be studied in children provided there is a pediatric disease that is like a non–orphan disease occurring in adults, and it is likely the new agent will be prescribed to children.[69,70] The FDA considers pcJIA the pediatric correlate of adult RA.[71] Likewise, cSLE is considered the pediatric correlate of SLE in adults. The FDA has recently recognized JPsA as the pediatric correlate of PsA and ERA as the pediatric counterpart of nr-axSpA. In 2012, the FDA Safety and Innovation Act (FDASIA)[72] made the PREA and BPCA permanent, which was an important development. The FDASIA requires earlier pediatric study plan (PSP) submission by drug manufacturers subject to the PREA and gives the FDA new authority to ensure PREA requirements are addressed in a timelier manner.

The case of extrapolation and global alignment of regulatory agencies. Integration of pediatric planning and exclusivity requests are now part of regular new drug and biologic product development programs at pharmaceutical companies. Key documents include the initial PSP for submission to, and approval by, the Center for Drug Evaluation and Research or the Center for Biologics Evaluation and Research at the FDA (Figure 1).[73,74] Although a PSP is only required for product development programs that are subject to the PREA, the FDA encourages sponsors to include all potential pediatric development plans for the product in the PSP, including those plans that may be studied under the BPCA. Such plans can form the basis of a

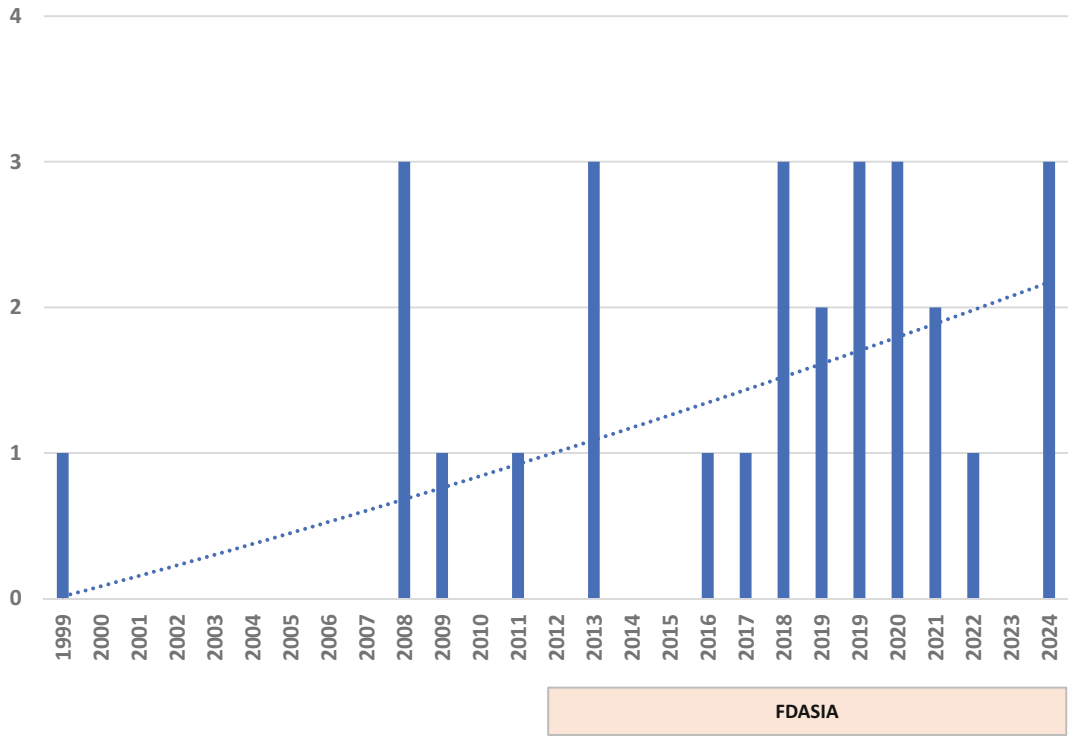


Figure 1. Biologic DMARD and targeted synthetic DMARD approvals by the US FDA for PRDs over time. DMARD approvals for PRDs have accelerated since the passage of the FDASIA. As of June 2024, already three medications have received regulatory approval for treatment of PRDs. The dotted line represents the trendline of new medication approvals by the FDA. DMARD, disease-modifying antirheumatic drug; FDA, Food and Drug Administration; FDASIA, FDA Safety and Innovation Act; PRD, pediatric rheumatic disease.

proposed pediatric study request (PPSR) that can be submitted to obtain a Written Request from the FDA. The FDA can issue a Written Request on its own initiative or at the request of an interested party. If a sponsor wishes to obtain pediatric exclusivity, the sponsor is strongly encouraged to submit a PPSR to expedite the FDA's issuance of a Written Request. Therefore, the same product may have both a PSP and a Written Request, thereby including both mandatory and voluntary studies. This could lead to the potential benefit of incentives of the BPCA if the submission is compliant with the studies contained in the Written Request.[75]

The EMA adopted rules for pediatric drug development in the European Union in 2013.[76] Importantly, the EMA also passed legislation concerning pediatric drug testing and approval. However, the mandate to perform pediatric studies and the provision of extended marketing exclusivity are governed by different processes than in the United States.

Both the EMA and FDA confirmed that the existence of pediatric research networks, such as the PRCSG, Pediatric International Trials Organisation (PRINTO), and CARRA, are of utmost importance to realize the potential of this pediatric legislation.[5] Like the PSP, a pediatric investigational plan (PIP) must be submitted to the Paediatric Committee (PDCO), the EMA scientific committee responsible for activities concerning medicine testing in children. The PDCO oversees the labeling of such medicines in the European Union.

One difference between US and European legislation is that PIP submission to the PDCO takes place by the end of phase I development in adults, whereas PSP submission to the FDA is expected to occur later, around the end of phase II development in adults. Differences in the FDA and EMA policies and/or disease assumptions have led to different mandates for study of PRD-associated disease manifestations. For example, the FDA only demands studies of JIA manifestations that also occur in adults. This is one of the reasons why studies in uveitis, which are generally absent in RA, are not required by the FDA. Conversely, the PIPs submitted to the EMA/PDCO usually require the study of JIA-associated uveitis.

The FDA and EMA, along with other global regulatory authorities and industry representatives, participated to add an addendum to the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use E11 guideline, "Clinical Investigation of Medicinal Products in the Pediatric Population."[77]

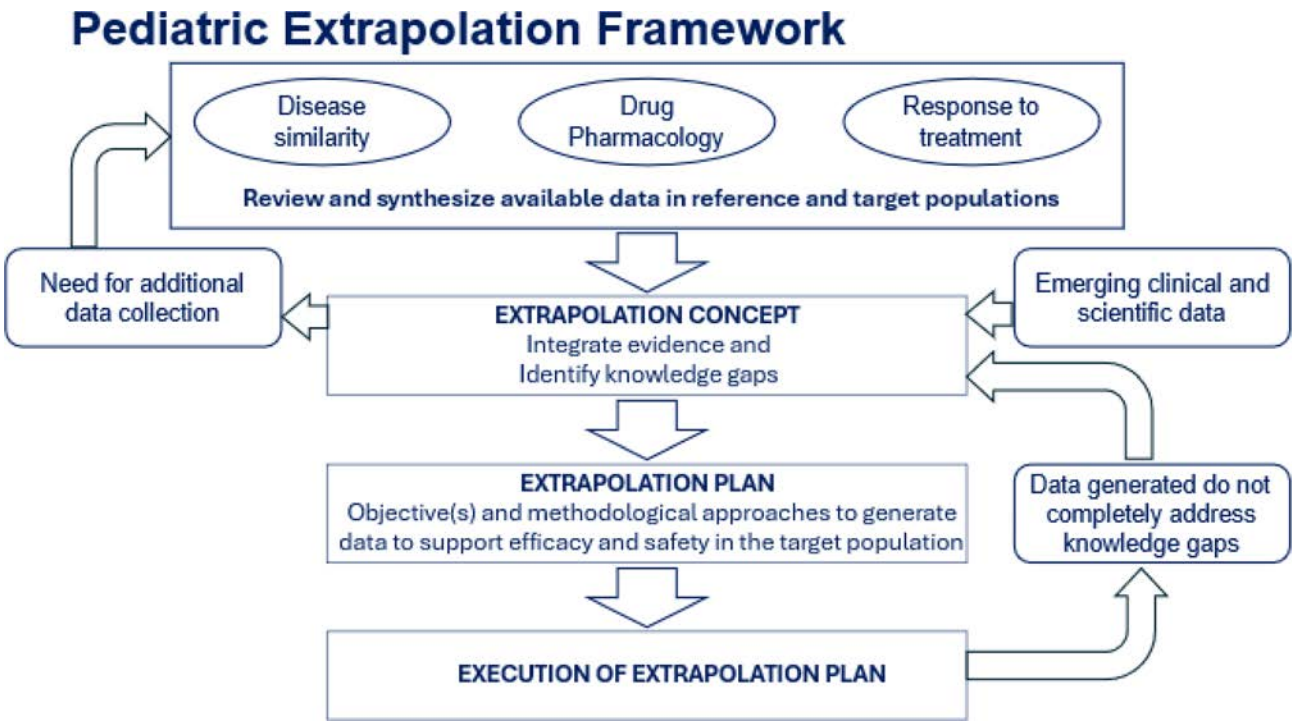


Figure 2. Pediatric extrapolation framework for new approvals for pediatric rheumatic diseases. The extrapolation framework consists of three parts: development of a pediatric extrapolation concept, the creation of a pediatric extrapolation plan, and execution of a pediatric extrapolation plan. The pediatric extrapolation concept is developed through comprehensive and detailed review of existing information about the range of factors that define the disease, the drug pharmacology, and the clinical response to treatment across the reference and target populations. Methods to review and synthesize these data can include quantitative approaches such as statistical methods, modeling, and simulation. Synthesis of the data should be conducted to both understand the strength of the known data and to identify important gaps in knowledge that will inform what additional data, if any, are required. The pediatric extrapolation plan should include the objective(s) and methodologic approaches for the data that need to be generated to support efficacy and safety in the target population for the purpose of regulatory decision-making. The existing pediatric extrapolation plan can be modified based on emerging relevant data to reflect current scientific and clinical understanding. *Source: ICH Harmonised Guideline: Pediatric Extrapolation E11A*.[74] Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25482/abstract.

This addendum is intended to address new scientific and technical knowledge advances in pediatric drug development. More recently, the FDA published draft guidance entitled "General Clinical Pharmacology Considerations for Pediatric Studies of Drugs, Including Biological Products Guidance for Industry"[69] and has expanded recommendations for the study of rare diseases.[70] Drug development for nonsystemic forms of JIA was the focus of the 2019 workshop of the FDA (Accelerating Drug Development for Polyarticular Juvenile Idiopathic Arthritis [pJIA]) in which clinical trial networks such as the PRCSG, CARRA, and other stakeholders discussed the opportunities of extrapolation in facilitating medication approvals in the United States.[78]

Impact of regulation on PRD medication approvals

Before 1999, a myriad of nonsteroidal anti-inflammatory medications, including aspirin and several nonsteroidal anti-inflammatory drugs, were evaluated and many were approved for the treatment of JIA. Methotrexate received approval for treating JIA in February 1999. In May 1999, etanercept received regulatory approval by the FDA for the treatment of pcJIA, one year after its approval in treating RA. There was a notable hiatus with no bDMARD approval for a PRD from 1999 to 2008, the latter year being when both adalimumab and abatacept were approved for treatment of pcJIA. Conversely, after the passage of FDASIA in 2012,[72] the time to biologic and drug approvals for the treatment of PRDs has quickened, with a total of 14 bDMARDs and tsDMARDs receiving an indication and a total of 24 medications now approved for the treatment of a PRD (Figure 2). Most medications have been approved for the treatment of pcJIA. However, the first medication approval for sJIA occurred in 2011, and for JPsA and ERA, the first bDMARD was approved in 2021. Notable other indications are for CAPS, starting in 2008, and other monogenic autoinflammatory diseases in 2016. The time between FDA approval of a bDMARD or a tsDMARD for treatment of an adult rheumatic disease and the corresponding PRD ranges between one year (etanercept) and eight to nine years (IV belimumab, ustekinumab). In rarer PRDs, approvals have occurred even before or concurrent with approval for adults.

Currently, there are more than 10 bDMARDs or tsDMARDs approved for adult forms of inflammatory arthritis (ie, RA, nr-axSpA, spondyloarthritis, and PsA) that are not also approved for the treatment of the corresponding PRDs.[71] This further underscores the unmet need for additional improvements in the path to approval of medications for PRD. There continues to be a poor prognosis of uveitis associated with PRDs, and there is only a single approved medication for uveitis, adalimumab, which carries a risk of anti-drug antibody formation that can lead to loss of efficacy.[79,80]

Conclusions

In summary, the licensure of bDMARDs and tsDMARDs for PRDs over the preceding 25 years has significantly improved the prognosis of JIA and select monogenic autoinflammatory diseases. In this context, trials in PRDs that fail to result in licensure must be considered as lost opportunities to potentially improve PRD outcomes. Research networks, such as the PRCSG, CARRA, and PRINTO, with their profound knowledge of PRDs and experience with PRD clinical trials and existing medical needs, continue to be central to supporting the approval of effective and safe medications in the years to come. Approved medications for PRDs have a well-defined safety and efficacy profile, and proper dosing has been established. FDA approval is often a prerequisite for medication coverage by insurance and health care authorities and, hence, access to these medications.

Specifically, there remains a medical need to advance approval of medications for PRD manifestations that occur most often in the pediatric forms of rheumatic diseases than in adults, such as uveitis. We also propose to carefully review the Orphan Drug Approval Law and its interpretation to avoid exclusion of pediatric patients when medication approvals for adult patients with rheumatic diseases are sought under this regulation in the future. It is worth noting that none of the advanced therapies used for the treatment of PRDs yield a cure or long-term remission of disease. Hence, additional bDMARDs and tsDMARDs are likely needed to control signs and symptoms of PRDs for years and possibly for life. Furthermore, discussions must occur with regulators to determine whether modern clinical pharmacology techniques and strategies could allow for the estimation of pediatric drug exposures and dosing regimens to allow bDMARDs that are currently only approved for the adult rheumatic disease to be approved for the corresponding PRD based on full extrapolation of efficacy. Ongoing discussions between regulators and the pediatric rheumatology community regarding ways to implement FDA-proposed strategies for the study of rare diseases in the PRD space are expected to improve strategies. Considering the improvement of JIA outcomes with bDMARD treatments, the noted existing burden of PRDs, and the noted delay between approval of adult rheumatic diseases and PRDs, accelerated FDA approval of additional bDMARDs and tsDMARDs for PRDs remains of major importance.

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 Shishov 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. Sacks JJ, Helmick CG, Luo YH, et al. Prevalence of and annual ambulatory health care visits for pediatric arthritis and other rheumatologic

conditions in the United States in 2001-2004. Arthritis Care Res (Hoboken) 2007;57:1439–1445.
2. Petty RE, Southwood TR, Baum J, et al. Revision of the proposed classification criteria for juvenile idiopathic arthritis: Durban, 1997. J Rheumatol 1998;25:1991–1994.
3. Silva CA, Avcin T, Brunner HI. Taxonomy for systemic lupus erythematosus with onset before adulthood. Arthritis Care Res (Hoboken) 2012;64:1787–1793.
4. 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. Ann Rheum Dis 2017;76:1955–1964.
5. Turner MA, Attar S, de Wildt SN, et al. Roles of clinical research networks in pediatric drug development. Clin Ther 2017;39:1939–1948.
6. Lovell DJ, Giannini EH, Reiff A, et al. Etanercept in children with polyarticular juvenile rheumatoid arthritis. Pediatric Rheumatology Collaborative Study Group. N Engl J Med 2000;342:763–769.
7. Giannini EH, Ilowite NT, Lovell DJ, et al. Long-term safety and effectiveness of etanercept in children with selected categories of juvenile idiopathic arthritis. Arthritis Rheum 2009;60:2794–2804.
8. Onel KB, Horton DB, Lovell DJ, et al. 2021 American College of Rheumatology guideline for the treatment of juvenile idiopathic arthritis: therapeutic approaches for oligoarthritis, temporomandibular joint arthritis, and systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2022;74:521–537.
9. Ruperto N, Lovell DJ, Cuttica R, et al. A randomized, placebo-controlled trial of infliximab plus methotrexate for the treatment of polyarticular-course juvenile rheumatoid arthritis. Arthritis Rheum 2007;56:3096–3106.
10. Brunner HI, Ruperto N, Tzaribachev N, et al. Subcutaneous golimumab for children with active polyarticular-course juvenile idiopathic arthritis: results of a multicentre, double-blind, randomised-withdrawal trial. Ann Rheum Dis 2018;77:21–29.
11. Singh R, Ivaturi VD, Penzenstadler J, et al. Response similarity assessment between polyarticular juvenile idiopathic arthritis and adult rheumatoid arthritis for biologics. Clin Pharmacol Ther 2021;110:98–107.
12. McNulty R. Sarilumab FDA approved for active polyarticular juvenile idiopathic arthritis. Am J Manage Care June 11, 2024. Accessed June 17, 2024. https://www.ajmc.com/view/sarilumab-fda-approved-for-active-polyarticular-juvenile-idiopathic-arthritis
13. Ramanan AV, Dick AD, Jones AP, et al. Adalimumab in combination with methotrexate for refractory uveitis associated with juvenile idiopathic arthritis: a RCT. Health Technol Assess 2019;23:1–140.
14. Nishimoto N, Yoshizaki K, Maeda K, et al. Toxicity, pharmacokinetics, and dose-finding study of repetitive treatment with the humanized anti-interleukin 6 receptor antibody MRA in rheumatoid arthritis. Phase I/II clinical study. J Rheumatol 2003;30:1426–1435.
15. Choy E. Clinical experience with inhibition of interleukin-6. Rheum Dis Clin North Am 2004;30:405–415, viii.
16. Brunner HI, Ruperto N, Ramanan AV, et al. Long-term efficacy and safety of subcutaneous tocilizumab in clinical trials of polyarticular or systemic juvenile idiopathic arthritis. Rheumatology (Oxford) 2024;63:2535–2546.
17. Brunner HI, Ruperto N, Zuber Z, et al. Efficacy and safety of tocilizumab in patients with polyarticular-course juvenile idiopathic arthritis: results from a phase 3, randomised, double-blind withdrawal trial. Ann Rheum Dis 2015;74:1110–1117.
18. Bharucha KN, Brunner HI, Calvo Penadés I, et al. Growth during tocilizumab therapy for polyarticular-course juvenile idiopathic arthritis: 2-year data from a phase III clinical trial. J Rheumatol 2018;45:1173–1179.
19. Malattia C, Ruperto N, Pederzoli S, et al. Tocilizumab may slow radiographic progression in patients with systemic or polyarticular-course juvenile idiopathic arthritis: post hoc radiographic analysis from two randomized controlled trials. Arthritis Res Ther 2020;22:211.
20. Lovell DJ, Tzaribachev N, Henrickson M, et al. Safety and effectiveness of abatacept in juvenile idiopathic arthritis: results from the PRINTO/PRCSG registry. Rheumatology (Oxford) 2024;63:SI195–SI206.
21. Brunner HI, Tzaribachev N, Vega-Cornejo G, et al. Subcutaneous abatacept in patients with polyarticular-course juvenile idiopathic arthritis: results from a phase III open-label study. Arthritis Rheumatol 2018;70:1144–1154.
22. Amarilyo G, Tarp S, Foeldvari I, et al. Biological agents in polyarticular juvenile idiopathic arthritis: a meta-analysis of randomized withdrawal trials. Semin Arthritis Rheum 2016;46:312–318.
23. Tarp S, Amarilyo G, Foeldvari I, et al. Efficacy and safety of biological agents for systemic juvenile idiopathic arthritis: a systematic review and meta-analysis of randomized trials. Rheumatology (Oxford) 2016;55:669–679.
24. Ruperto N, Brunner HI, Quartier P, et al. Canakinumab in patients with systemic juvenile idiopathic arthritis and active systemic features: results from the 5-year long-term extension of the phase III pivotal trials. Ann Rheum Dis 2018;77:1710–1719.
25. Yokota S, Imagawa T, Mori M, et al. Long-term treatment of systemic juvenile idiopathic arthritis with tocilizumab: results of an open-label extension study in Japan. Ann Rheum Dis 2013;72:627–628.
26. De Benedetti F, Brunner HI, Ruperto N, et al. Randomized trial of tocilizumab in systemic juvenile idiopathic arthritis. N Engl J Med 2012;367:2385–2395.
27. Ruperto N, Brunner HI, Quartier P, et al. Two randomized trials of canakinumab in systemic juvenile idiopathic arthritis. N Engl J Med 2012;367:2396–2406.
28. Ruperto N, Quartier P, Wulffraat N, et al. A phase II, multicenter, open-label study evaluating dosing and preliminary safety and efficacy of canakinumab in systemic juvenile idiopathic arthritis with active systemic features. Arthritis Rheum 2012;64:557–567.
29. Quartier P, Allantaz F, Cimaz R, et al. A multicentre, randomised, double-blind, placebo-controlled trial with the interleukin-1 receptor antagonist anakinra in patients with systemic-onset juvenile idiopathic arthritis (ANAJIS trial). Ann Rheum Dis 2011;70:747–754.
30. Ter Haar NM, van Dijkhuizen EHP, Swart JF, et al. Treatment to target using recombinant interleukin-1 receptor antagonist as first-line monotherapy in new-onset systemic juvenile idiopathic arthritis: results from a five-year follow-up study. Arthritis Rheumatol 2019;71:1163–1173.
31. Vastert SJ, de Jager W, Noordman BJ, et al. Effectiveness of first-line treatment with recombinant interleukin-1 receptor antagonist in steroid-naive patients with new-onset systemic juvenile idiopathic arthritis: results of a prospective cohort study. Arthritis Rheumatol 2014;66:1034–1043.
32. Lovell DJ, Giannini EH, Reiff AO, et al. Long-term safety and efficacy of rilonacept in patients with systemic juvenile idiopathic arthritis. Arthritis Rheum 2013;65:2486–2496.
33. Ilowite NT, Prather K, Lokhnygina Y, et al. Randomized, double-blind, placebo-controlled trial of the efficacy and safety of rilonacept in the treatment of systemic juvenile idiopathic arthritis. Arthritis Rheumatol 2014;66:2570–2579.
34. Grom AA, Ilowite NT, Pascual V, et al. Rate and clinical presentation of macrophage activation syndrome in patients with systemic juvenile idiopathic arthritis treated with canakinumab. Arthritis Rheumatol 2016;68:218–228.
35. Schulert GS, Grom AA. Pathogenesis of macrophage activation syndrome and potential for cytokine-directed therapies. Annu Rev Med 2015;66:145–159.

36. Binstadt BA, Nigrovic PA. The conundrum of lung disease and drug hypersensitivity-like reactions in systemic juvenile idiopathic arthritis. Arthritis Rheumatol 2022;74:1122–1131.
37. Adult-related conditions that qualify for a waiver because they rarely or never occur in pediatrics. Food and Drug Administration. September 30, 2024. Accessed January 20, 2025. https://www.fda.gov/media/101440/download 2023
38. Leu JH, Shiff NJ, Clark M, et al. Intravenous golimumab in patients with polyarticular juvenile idiopathic arthritis and juvenile psoriatic arthritis and subcutaneous ustekinumab in patients with juvenile psoriatic arthritis: extrapolation of data from studies in adults and adjacent pediatric populations. Paediatr Drugs 2022;24:699–714.
39. Ruperto N, Brunner HI, Pacheco-Tena C, et al. Open-label phase 3 study of intravenous golimumab in patients with polyarticular juvenile idiopathic arthritis. Rheumatology (Oxford) 2021;60:4495–4507.
40. Brunner HI, Ruperto N, Tzaribachev N, et al. Subcutaneous golimumab for children with active polyarticular-course juvenile idiopathic arthritis: results of a multicentre, double-blind, randomised-withdrawal trial. Ann Rheum Dis 2018;77:21–29.
41. Brunner HI, Foeldvari I, Alexeeva E, et al. Secukinumab in enthesitis-related arthritis and juvenile psoriatic arthritis: a randomised, double-blind, placebo-controlled, treatment withdrawal, phase 3 trial. Ann Rheum Dis 2023;82:154–160.
42. Ringold S, Angeles-Han ST, Beukelman T, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the treatment of juvenile idiopathic arthritis: therapeutic approaches for non-systemic polyarthritis, sacroiliitis, and enthesitis. Arthritis Care Res (Hoboken) 2019;71:717–734.
43. Weiss PF, Fuhlbrigge RC, von Scheven E, et al. Children with enthesitis-related arthritis and possible benefits from treatments for adults with spondyloarthritis. Arthritis Care Res (Hoboken) 2022;74:1058–1064.
44. Ramanan AV, Quartier P, Okamoto N, et al. Baricitinib in juvenile idiopathic arthritis: an international, phase 3, randomised, double-blind, placebo-controlled, withdrawal, efficacy, and safety trial. Lancet 2023;402:555–570.
45. Ruperto N, Brunner HI, Synoverska O, et al. Tofacitinib in juvenile idiopathic arthritis: a double-blind, placebo-controlled, withdrawal phase 3 randomised trial. Lancet 2021;398:1984–1996.
46. Brunner HI, Horneff G, Foeldvari I, et al. Safety and efficacy of upadacitinib for pediatric patients with polyarticular course juvenile idiopathic arthritis: an interim analysis of an open-label, phase 1 trial. Ann Rheum Dis 2023;82:108–109.
47. Brunner HI, Abud-Mendoza C, Viola DO, et al. Safety and efficacy of intravenous belimumab in children with systemic lupus erythematosus: results from a randomised, placebo-controlled trial. Ann Rheum Dis 2020;79:1340–1348.
48. Balevic SJ, Niu J, Chen J, et al. Extrapolation of adult efficacy data to pediatric systemic lupus erythematosus: evaluating similarities in exposure–response. J Clin Pharmacol 2023;63:105–118.
49. Brunner HI, Oscar VD, Dimelow R, et al. Pharmacokinetic, pharmacodynamic, and safety profile of subcutaneous belimumab in pediatric patients with systemic lupus erythematosus: analysis of data from a multicenter, open-label trial. *Arthritis* Rheumatol 2023;75(suppl 9).
50. Goldbach-Mansky R, Shroff SD, Wilson M, et al. A pilot study to evaluate the safety and efficacy of the long-acting interleukin-1 inhibitor rilonacept (interleukin-1 Trap) in patients with familial cold autoinflammatory syndrome. Arthritis Rheum 2008;58:2432–2442.
51. Kuemmerle-Deschner JB, Gautam R, George AT, et al. A systematic literature review of efficacy, effectiveness and safety of biologic therapies for treatment of familial Mediterranean fever. Rheumatology (Oxford) 2020;59:2711–2724.
52. Hashkes PJ, Spalding SJ, Giannini EH, et al. Rilonacept for colchicine-resistant or -intolerant familial Mediterranean fever: a randomized trial. Ann Intern Med 2012;157:533–541.
53. De Benedetti F, Gattorno M, Anton J, et al. Canakinumab for the treatment of autoinflammatory recurrent fever syndromes. N Engl J Med 2018;378:1908–1919.
54. Oddis CV, Reed AM, Aggarwal R, et al. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. Arthritis Rheum 2013;65:314–324.
55. Aggarwal R, Charles-Schoeman C, Schessl J, et al. Trial of intravenous immune globulin in dermatomyositis. N Engl J Med 2022;387:1264–1278.
56. Palmisani E, Solari N, Magni-Manzoni S, et al. Correlation between juvenile idiopathic arthritis activity and damage measures in early, advanced, and longstanding disease. Arthritis Rheum 2006;55:843–849.
57. Michels H, Häfner R, Morhart R, et al. Five year follow-up of a prospective cohort of juvenile chronic arthritis with recent onset. Clin Rheumatol 1987;6(Suppl 2):87–92.
58. Oen K, Malleson PN, Cabral DA, et al. Disease course and outcome of juvenile rheumatoid arthritis in a multicenter cohort. J Rheumatol 2002;29:1989–1999.
59. Walco GA, Varni JW, Ilowite NT. Cognitive-behavioral pain management in children with juvenile rheumatoid arthritis. Pediatrics 1992;89:1075–1079.
60. Milatz F, Hansmann S, Klotsche J, et al. Level and correlates of physical activity among children and adolescents with juvenile idiopathic arthritis compared to controls: results from a German nationwide prospective observational cohort study. Pediatr Rheumatol Online J 2024;22:39.
61. Bingham CA, Harris JG, Qiu T, et al. Pediatric Rheumatology Care and Outcomes Improvement Network's quality measure set to improve care of children with juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2023;75:2442–2452.
62. Brunner HI, Schanberg LE, Kimura Y, et al. New medications are needed for children with juvenile idiopathic arthritis. Arthritis Rheumatol 2020;72:1945–1951.
63. Grazziotin LR, Currie G, Twilt M, et al. Real-world data reveals the complexity of disease modifying anti-rheumatic drug treatment patterns in juvenile idiopathic arthritis: an observational study. Pediatr Rheumatol Online J 2022;20:25.
64. Asbury CH. The Orphan Drug Act. The first 7 years. JAMA 1991;265:893–897.
65. Mikami K. Orphans in the market: the history of orphan drug policy. Soc Hist Med 2019;32:609–630.
66. Inflation Reduction Act of 2022. Internal Revenue Service. Updated November 19, 2024. Accessed June 10, 2024. https://www.irs.gov/inflation-reduction-act-of-2022
67. National Institute of Child Health and Human Development. Best Pharmaceuticals for Children Act. National Institute of Child Health and Human Development; 2002. Accessed June 3, 2024. https://www.congress.gov/107/plaws/publ109/PLAW-107publ109.pdf
68. Food and Drug Administration. Pediatric Research Equity Act of 2003. Food and Drug Administration; 2003. Accessed June 3, 2024. https://www.congress.gov/108/plaws/publ155/PLAW-108publ155.pdf
69. Christensen ML. Best Pharmaceuticals for Children Act and Pediatric Research Equity Act: time for permanent status. J Pediatr Pharmacol Ther 2012;17:140–141.
70. Califf RM. Best Pharmaceuticals for Children Act and Pediatric Research Equity Act. Status Report to Congress. Department of Health and Human Services, Food and Drug Administration;

2016. Accessed June 3, 2024. https://www.fda.gov/media/99184/download

71. Dunne J, Rodriguez WJ, Murphy MD, et al. Extrapolation of adult data and other data in pediatric drug-development programs. Pediatrics 2011;128:e1242–e1249.
72. Food and Drug Administration. Food and Drug Administration Safety and Innovation Act. Food and Drug Administration; 2012. Accessed June 3, 2024. https://www.govinfo.gov/content/pkg/PLAW-112publ144/pdf/PLAW-112publ144.pdf
73. Food and Drug Administration. Draft Guidance for Industry on Pediatric Study Plans: Content of and Process for Submitting Initial Pediatric Study Plans and Amended Pediatric Study Plans; Availability. Food and Drug Administration; 2013. Accessed June 3, 2024. https://www.regulations.gov/document/FDA-2013-D-0814-0001
74. Food and Drug Administration. ICH Harmonised Guideline: Pediatric Extrapolation E11A. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; 2022. Accessed June 3, 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/e11a-pediatric-extrapolation
75. Penkov D, Tomasi P, Eichler I, et al. Pediatric medicine development: an overview and comparison of regulatory processes in the European Union and United States. Ther Innov Regul Sci 2017;51:360–371.
76. Committee for Medicinal Products for Human Use; Paediatric Committee. *Guideline on Pharmaceutical Development of Medicines for Paediatric Use*. European Medicines Agency; 2013. Accessed June 3, 2024. https://www.tga.gov.au/sites/default/files/2024-07/Guidelineonpharmaceuticaldevelopmentofmedicinesforpaediatric%20use%20.pdf
77. Food and Drug Administration. E11(R1) Addendum: Clinical Investigation of Medicinal Products in the Pediatric Population; International Council for Harmonisation; Guidance for Industry. Food and Drug Administration; 2018. Accessed June 3, 2024. https://www.regulations.gov/document/FDA-2016-D-3848-0005
78. Schanberg LE, Mulugeta LY, Akinlade B, et al. Therapeutic development in polyarticular course juvenile idiopathic arthritis: extrapolation, dose selection, and clinical trial design. Arthritis Rheumatol 2023;75:1856–1866.
79. Miraldi Utz V, Angeles-Han ST, Mwase N, et al. Alternative biologic therapy in children failing conventional TNFα inhibitors for refractory, noninfectious, chronic anterior uveitis. Am J Ophthalmol 2022;244:183–195.
80. Cecchin V, Zannin ME, Ferrari D, et al. Long-term safety and efficacy of adalimumab and infliximab for uveitis associated with juvenile idiopathic arthritis. J Rheumatol 2018;45:1167–1172.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 584–593
DOI 10.1002/acr.25465

American College of Rheumatology
Empowering Rheumatology Professionals

Pharmacokinetics, Efficacy, and Safety of Upadacitinib in Pediatric Patients With Polyarticular-Course Juvenile Idiopathic Arthritis: An Interim Analysis of an Open-Label, Phase 1 Trial

Hermine I. Brunner,[1] Anna Shmagel,[2] Gerd Horneff,[3] Ivan Foeldvari,[4] Jordi Antón,[5] Athimalaipet V. Ramanan,[6] Yuli Qian,[2] Kristina Unnebrink,[7] Shuai Hao,[2] Heidi S. Camp,[2] Nasser Khan,[2] Wei Liu,[2] and Mohamed-Eslam F. Mohamed[2]

Objective. This work aimed to evaluate the pharmacokinetics, efficacy, and safety of upadacitinib, an oral selective JAK inhibitor, in pediatric patients with polyarticular-course juvenile idiopathic arthritis (pcJIA).

Methods. In an open-label, phase 1 study (SELECT-YOUTH), enrolled patients, aged 2 to <18 years with pcJIA, received body weight–based upadacitinib doses using a twice-daily oral solution or once-daily extended-release tablet based on their body weight and ability to swallow tablets. The study included a 7-day pharmacokinetic assessment, followed by a long-term efficacy and safety evaluation for up to 156 weeks, including an additional long-term safety cohort. This interim analysis included available pharmacokinetic and safety data and efficacy data collected through week 48.

Results. A total of 57 patients received upadacitinib. The median time to maximum upadacitinib concentration was approximately three hours and one hour for the tablet and oral solution regimens, respectively; the harmonic mean functional half-life was approximately five hours and two hours, respectively. Juvenile idiopathic arthritis American College of Rheumatology 30, 50, 70, 90, and 100 responses at week 12 were 91.8%, 89.8%, 69.4%, 49.0%, and 32.7%, respectively. Efficacy was generally maintained through week 48, and improvement in additional efficacy end points was also observed. At a median exposure duration of 412 days, 52 of 57 patients reported adverse events; of these, 6 experienced serious adverse events. Adverse events were predominately mild to moderate in severity and consistent with the known safety profile of upadacitinib.

Conclusion. This interim analysis demonstrates that the bodyweight-based dosing regimen of upadacitinib was well tolerated and efficacious in pediatric patients with pcJIA.

INTRODUCTION

Juvenile idiopathic arthritis (JIA) is an umbrella term for chronic arthritis of unknown etiology beginning before the age of 16 years. Several JIA subtypes, per International League of Associations for Rheumatology criteria,[1] are classified as polyarticular-course juvenile idiopathic arthritis (pcJIA), including rheumatoid factor (RF)–positive and RF-negative polyarthritis, extended oligoarthritis, and systemic JIA (sJIA) without systemic symptoms. pcJIA is characterized by a high rate of joint and tissue damage compared with the other subtypes of JIA and accounts for 30% to 50% of all patients with JIA.[2–4] Substantial functional impairment, risk of growth disturbances, joint deformity, and long-term disability are seen across JIA subtypes, including pcJIA.[3] Despite

ClinicalTrials.gov identifier: NCT03725007.
EudraCT database no.: 2018-000715-25.
Supported by AbbVie.
[1]Hermine I. Brunner, MD: University of Cincinnati, Cincinnati Children's Hospital Medical Center, Ohio; [2]Anna Shmagel, MD, Yuli Qian, PhD, Shuai Hao, PhD, Heidi S. Camp, PhD, Nasser Khan, MD, Wei Liu, PhD, Mohamed-Eslam F. Mohamed, PhD: AbbVie Inc., North Chicago, Illinois; [3]Gerd Horneff, MD: Asklepios Clinic Sankt Augustin, Sankt Augustin, and University Hospital of Cologne, Cologne, Germany; [4]Ivan Foeldvari, MD: Hamburger Zentrum für Kinder und Jugendrheumatologie, am Klinikum Eilbek, Hamburg, Germany; [5]Jordi Antón, MD: Hospital Sant Joan de Déu, Universitat de Barcelona, Barcelona, Spain; [6]Athimalaipet V. Ramanan, MD: Bristol Royal Hospital for Children & Translational Health Sciences, University of Bristol, Bristol, United Kingdom; [7]Kristina Unnebrink, PhD: AbbVie Deutschland GmbH & Co. KG, Ludwigshafen, Germany.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25465).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25465.

Address correspondence via email to Mohamed-Eslam F. Mohamed, PhD, at mohamed-eslam.mohamed@abbvie.com.

Submitted for publication May 21, 2024; accepted in revised form October 29, 2024.

SIGNIFICANCE & INNOVATIONS

- This study reports the first characterization of upadacitinib pharmacokinetics, efficacy, and safety in patients with polyarticular-course juvenile idiopathic arthritis.
- An adaptive dose selection feature was implemented in this study to achieve target upadacitinib exposures in pediatric patients.
- Data from this study provided essential support for an innovative pediatric extrapolation approach that led to the regulatory approval of upadacitinib for polyarticular juvenile idiopathic arthritis in the United States.

the approval of biologic disease-modifying antirheumatic drugs (bDMARDs), currently approved medications do not suffice to yield clinical remission in the majority of patients who are also prone to develop intolerance or lose response over time.[5] With >45% of patients with pcJIA potentially having active disease in adulthood,[6] additional novel treatment options are warranted to treat JIA, including oral formulations, the preferred route of administration in the pediatric population.

Upadacitinib is an oral JAK inhibitor with proven efficacy in adult rheumatoid arthritis (RA), psoriatic arthritis, atopic dermatitis, axial spondylarthritis, ulcerative colitis, Crohn disease, and in adolescent atopic dermatitis in patients 12 years of age or older.[7] Upadacitinib potently inhibits JAK1 but is less potent against JAK2, JAK3, and tyrosine kinase 2 (Tyk2), with downstream inhibition of several proinflammatory cytokine signals. Clinical efficacy through inhibition of the JAK/STAT pathway in this population has recently been demonstrated with other JAK inhibitors[8–10]; thus, therapeutic benefit of upadacitinib is anticipated in pcJIA.

Here, we present the interim results from a phase 1 study (SELECT-YOUTH) designed with the objectives of evaluating the pharmacokinetics, efficacy, and safety of upadacitinib in patients with pcJIA aged 2 to <18 years. This is the first clinical report of upadacitinib in patients with pcJIA.

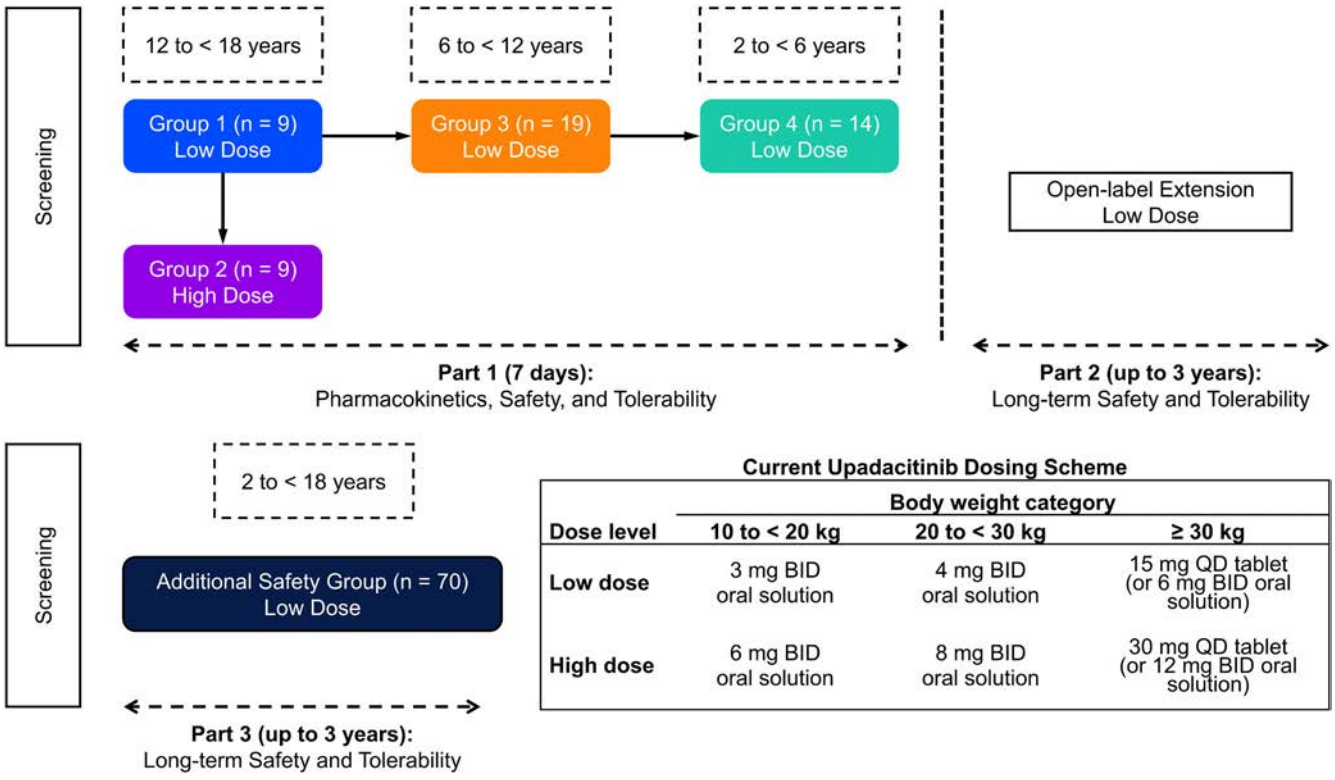
PATIENTS AND METHODS

Study design. This is an ongoing, phase 1, open-label, nonrandomized, multiple-dose study with a target enrollment of ~124 patients, designed to evaluate the pharmacokinetics, efficacy, safety, and tolerability of upadacitinib in pediatric patients (2 to <18 years old) with pcJIA. This is an interim analysis based on a data cutoff date of September 22, 2022, where the seven-day multiple-dose pharmacokinetic evaluation in part 1 had been completed, and extended safety and efficacy assessments in parts 2 and 3 are ongoing. (For information on data sharing, see Supplementary Materials).

A study diagram is shown in Figure 1. As of the interim cutoff date, 57 pediatric patients with pcJIA were enrolled across 14 sites in the United States, Germany, Hungary, Puerto Rico, and Spain. This study was conducted in accordance with the International Council for Harmonisation guidelines, applicable regulations, and guidelines governing clinical study conduct and the ethical principles that have their origin in the Declaration of Helsinki. The study protocol was approved by the Institutional Review Boards/Ethics Committee boards at each of the study sites, and all the participants' legally authorized representatives provided written informed consent before participation in the study. Patients were included in all discussions to obtain verbal and/or written assent.

Part 1 was an open-label, seven-day multiple-dose study that mainly evaluated the pharmacokinetics of upadacitinib in pediatric patients with pcJIA and was designed to allow for dose adjustment, if needed, based on available pharmacokinetic and clinical data. Patients from three age groups (12 to <18 years, 6 to <12 years, and 2 to <6 years) were enrolled in a staggered approach and received either low or high doses of upadacitinib (Figure 1). All patients who completed part 1 and were benefiting from study drug, per investigator judgment, with no ongoing adverse events (AEs) of special interest (AESIs) or serious adverse events (SAEs), based on investigator's clinical judgment and with patient/family's agreement, had the option to enroll in part 2. Part 2 was an open-label extension to evaluate the long-term safety and tolerability of upadacitinib at the low-dose level. Part 3 was an additional safety cohort including patients from all age groups (2 to <18 years) and was added to further evaluate the long-term safety and tolerability of the low dose of upadacitinib. Part 3 followed the same visit schedule as in part 2 but without intensive pharmacokinetic sampling. Patients in part 2 and part 3 were followed up to week 156, with an option to continue upadacitinib treatment beyond week 156 for patients who benefited from treatment. Descriptive efficacy was collected throughout all parts of the study.

Patients. Male and female pediatric patients aged 2 to <18 years and weighing ≥10 kg with active pcJIA in ≥5 joints (using the American College of Rheumatology [ACR] definition of active joint [presence of swelling not due to deformity or presence of limited motion with pain or tenderness])[11] at the time of screening were eligible to participate in the study. Patients needed not to have been diagnosed with enthesitis-related arthritis or juvenile psoriatic arthritis (PsA). Permitted prior and facultatively concomitant therapy included stable doses of nonsteroidal anti-inflammatory drugs (NSAIDs), low-dose glucocorticoids (≤0.2 mg/kg/day prednisone; daily maximum, 10 mg), and methotrexate (MTX; ≤20 mg/m^2 body surface area/week). Changes in the doses, initiation, and discontinuation of glucocorticoids, MTX, and NSAIDs were allowed during the study after the initial



Current Upadacitinib Dosing Scheme

	Body weight category		
Dose level	**10 to < 20 kg**	**20 to < 30 kg**	**≥ 30 kg**
Low dose	3 mg BID oral solution	4 mg BID oral solution	15 mg QD tablet (or 6 mg BID oral solution)
High dose	6 mg BID oral solution	8 mg BID oral solution	30 mg QD tablet (or 12 mg BID oral solution)

Figure 1. Study design diagram and dosing regimens. Some patients who received the original doses in part 1 had dose adjustments in part 2 (because of dose revision). Patients in the ≥30 kg body weight category were given the option to receive the oral solution formulation if unable to swallow the tablet. See Table S1 for original dosing regimens. BID, twice daily; QD, once daily.

pharmacokinetic assessment on study day 7, per local standard of care and at investigator's discretion. Prior bDMARDs had to have been discontinued before study day 1 per specified washout periods and were prohibited during the study. No prior exposure to JAK inhibitors was permitted.

Study medication: dosing and formulation of upadacitinib. Patients received either an immediate-release (IR) twice-daily oral solution formulation or an extended-release (ER) once-daily tablet formulation based on their body weight and ability to swallow tablets. The original dosing regimen was based on simulations using a previously developed population pharmacokinetic model in adults and assuming typical allometric scaling values of relevant pharmacokinetic parameters (ie, clearance and volume of distribution).[12] The low- and high-dose levels were selected to provide comparable plasma exposures in pediatric patients to the 15 mg and 30 mg once-daily ER tablets in adult patients with RA, respectively. The original dosing regimens are shown in Table S1. During the enrollment of group 3 in December 2020, a preplanned analysis was conducted to evaluate the available pharmacokinetic data collected from this study and another phase 1 study in pediatric patients with atopic dermatitis, which prompted revision of the original doses and body weight categories. The revised dosing regimen is provided in Figure 1.

Pharmacokinetic assessments. Blood samples for measuring upadacitinib plasma concentrations were collected into K_2 EDTA-containing collection tubes in part 1 on day 7 for once-daily tablet regimens before dosing (0 hour) and at 0.5, 1, 2, 3, 4, 6, 9, 12, and 24 hours after dosing and on day 7 for twice-daily oral solution regimens before dosing (0 hour) and at 0.25, 0.5, 1, 2, 4, 8, and 12 hours after dosing. Plasma concentrations of upadacitinib were determined using a validated liquid chromatography method with tandem mass spectrometric detection.[13] The lower limit of quantitation for upadacitinib was 0.05 ng/mL. Samples quantified below the lowest standard were reported as zero.

Pharmacokinetic analysis of upadacitinib was performed using noncompartmental methods in the Phoenix software (Certara, version 8.3.4). The pharmacokinetic parameters determined for upadacitinib were maximum plasma concentration (C_{max}), time to maximum observed plasma concentration (T_{max}), area under the plasma concentration versus time curve during a dosing interval (AUC_{tau}) at steady state, apparent oral clearance at steady state (CL_{ss}/F), and functional half-life ($t_{1/2}$). C_{max} and T_{max} were determined directly from the plasma concentration-

time data. Functional $t_{1/2}$ was calculated as ln (2)/(ln(C_{max}/C_{trough})/tau), where tau is the dosing interval and C_{trough} is the observed upadacitinib plasma concentration before the next dose. AUC_{tau} was calculated by the linear trapezoidal rule and represents $AUC_{0–24}$ for once-daily regimens and $AUC_{0–12}$ for twice-daily regimens. For twice-daily regimens, $AUC_{0–24}$ was calculated by multiplying $AUC_{0–12}$ by two. CL_{ss}/F, where F is the bioavailability, was calculated by dividing the administered dose by AUC_{tau}. In addition, bioavailability-adjusted CL_{ss}/F (Adjusted CL_{ss}/F) was calculated by multiplying CL_{ss}/F by 0.684 for once-daily regimens to account for the difference in bioavailability between the ER tablet and IR solution formulations. For twice-daily regimens, adjusted CL_{ss}/F was reported as CL_{ss}/F without a change in value.

Efficacy assessments. Efficacy end points were reported as observed by age group and overall, and descriptive analyses were performed for parts 1 and 2 at week 12 and through week 48. Efficacy end points included JIA ACR 30/50/70/90/100 responses, change from baseline in Juvenile Arthritis Disease Activity Score (JADAS) 27 based on measurement of C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), and change from baseline in individual JIA ACR and JADAS components: total number of active joints, defined as joints with swelling not due to deformity or joints with limitation of movement (LOM) with pain, tenderness, or both; number of joints with LOM; Childhood Health Assessment Questionnaire (C-HAQ); Physician's Global Assessment of Disease Activity (visual analog scale [VAS]); Patient's/Caregiver's Global Assessment of Overall Well-Being (VAS); ESR; and CRP.

Safety assessments. Safety evaluations performed during the study included incidence of treatment-emergent AEs (TEAEs), physical examination results, and clinical laboratory testing (hematology and chemistry). An AE with an onset date after the first dose of study drug and no more than 30 days after the last dose of study drug was defined as a TEAE. AEs were summarized by system organ class and preferred terms according to the Medical Dictionary for Regulatory Activities (MedDRA) (version 25.0). Patients reporting more than one AE for a given MedDRA preferred term were counted only once for that term. For assessing the severity of AEs, investigators classified TEAEs according to the NCI CTCAE version 5.0. AESIs were selected based on safety information reported for other JAK inhibitor products and upadacitinib data from preclinical studies and the adult RA development program. The AESI categories were identified per Standardised MedDRA Queries and Custom MedDRA Queries (SMQs/CMQs) or by MedDRA Preferred Terms.

Statistical analyses. Pharmacokinetic parameters of upadacitinib, determined based on data collected on day 7 in part 1, were summarized descriptively by age group, dose level, and formulation. Sample size was based on pharmacokinetic requirements in which complete data from 9 patients per dose level in age group 12 to <18 years and 18 patients in each of the age groups 6 to <12 years or 2 to <6 years would yield a 95% confidence interval within 60% and 140% of the geometric mean estimates of CL/F with at least 84.6% power, assuming 37% coefficient of variation for CL/F variability. The sample size in the additional safety cohort added in a protocol amendment was not based on statistical considerations, but on the need to collect additional long-term safety data for upadacitinib in pcJIA. Efficacy end points were reported using observed data (ie, without imputation of missing values). All end points are presented with descriptive statistics for overall patients and age groups at week 12 and through week 48. All patients enrolled into parts 1 and 2 who received at least one dose of upadacitinib were included in the efficacy analyses. For efficacy analyses, patients enrolled into part 3 were not included in the analyses because of their limited time on the study at the time of the data cutoff; for patients enrolled in parts 1 and 2, efficacy results were presented up to week 48 through the data cutoff date of September 22, 2022. All patients who received at least one dose of upadacitinib were included in the safety analyses. Safety data were summarized descriptively for overall patients and by age groups using all available safety data through the data cutoff date.

RESULTS

Patients and disposition. Enrollment for this study began July 4, 2019, and is ongoing. As of the interim data cutoff date of September 22, 2022, part 1 had been completed (n = 51) with all patients continuing to part 2, and six additional patients were enrolled in part 3. Part 2 and part 3 are ongoing. Among the 57 patients enrolled, 8 patients (14.0%) discontinued the study (see Figure S1). A summary of the patient demographics and disease characteristics for the study population is provided in Table 1 (see also Table S2 for data by group). The majority of enrolled patients were female (45/57, 78.9%) and had RF-negative polyarticular JIA (42/57, 73.7%). The enrolled patients were between 2 and 17 years old, with a mean disease duration of 3 years. At baseline, 23 (40.4%) patients received methotrexate, 11 (19.3%) patients received oral glucocorticoids, and 14 (24.6%) patients reported prior use of bDMARDs.

Pharmacokinetics. The mean plasma concentration-time profiles are presented in Figure 2 by dose level and formulations compared to adult reference profiles simulated based on pharmacokinetic analyses of upadacitinib data in phase 3 RA studies.[12] A summary of the pharmacokinetic parameters of upadacitinib at steady state (ie, on day 7) after administration of upadacitinib in patients enrolled in part 1 is provided in Table 2. During the pharmacokinetic assessment, 13 patients in group 3 and 12 patients in group 4 received revised doses (Figure 1), and all other patients received the original doses (Table S1). Upadacitinib C_{max} was reached within approximately three hours and one hour after the administration of the ER tablet and the IR oral solution formulations,

Table 1. Baseline demographics and disease characteristics of the participants*

Demographics	All patients (N = 57)
Female, n (%)	45 (78.9)
Age, years, mean ± SD (range)	9.5 ± 4.41 (2–17)
White, n (%)	55 (96.5)
Weight, kg, mean ± SD (range)	38.05 ± 20.380 (11.0–92.9)
pcJIA type, n (%)	
Extended oligoarticular JIA	7 (12.3)
RF-negative polyarticular JIA	42 (73.7)
RF-positive polyarticular JIA	7 (12.3)
Systemic JIA with active arthritis and without active systemic features	1 (1.8)
Duration of pcJIA diagnosis, y, mean ± SD (range)	2.959 ± 3.3387 (0.05–13.23)
Active joints, mean ± SD (range)	11.1 ± 7.74 (5–48)
C-HAQ	1.047 ± 0.7746 (0.00–2.88)
CRP (mg/L)	9.11 ± 18.958 (0.2–92.3)
ESR[a]	19.5 ± 19.02 (2–94)
Oral glucocorticoids, n (%)	11 (19.3)
Methotrexate, n (%)	23 (40.4)
Prior exposure to bDMARDs[b]	14 (24.6)

* bDMARD, biologic disease-modifying antirheumatic drug; C-HAQ, Childhood Health Assessment Questionnaire; CRP, C-reactive protein (normal range ≤2.87 mg/L); ESR, erythrocyte sedimentation rate (normal range 3–15 mm/h); JIA, juvenile idiopathic arthritis; pcJIA, polyarticular-course juvenile idiopathic arthritis; RF, rheumatoid factor.
[a] n = 53.
[b] Of these 14 patients, 10 patients were exposed to ≥3 prior bDMARDs, 3 patients were exposed to 2 prior bDMARDs, and 1 patient was exposed to 1 prior bDMARD. Prior bDMARDs included tumor necrosis factor inhibitors, abatacept, and tocilizumab.

respectively. Upadacitinib functional $t_{1/2}$ was approximately five hours within a once-daily dosing interval for the ER tablet and two hours within a twice-daily dosing interval for the IR solution.

Efficacy. Up to the data cutoff date, 50 of 51 patients from parts 1 and 2 had efficacy results available at week 12, and 37 patients had efficacy results available through week 48. A summary of the JIA ACR 30/70 responses, C-HAQ, and JADAS 27-CRP at week 12 is presented in Figure 3. Overall, the percentages of patients in parts 1 and 2 who achieved JIA ACR 30, 50, 70, 90, and 100 response at week 12 were 91.8%, 89.8%, 69.4%, 49.0%, and 32.7%, respectively. The two younger age groups (2 to <6 years and 6 to <12 years) had similar improvements in JIA ACR responses, and the oldest age group (12 to <18 years) had numerically lower rates of JIA ACR responses compared with the younger age groups without statistical significance. Response to upadacitinib was rapid, with 61.2% of patients achieving JIA ACR 30 as early as week 1. JIA ACR responses continued to improve at week 24 and were generally maintained through week 48 (Figure S2). Additional key efficacy measurements, including C-HAQ, total number of active joints, Physician's Global Assessment of Disease Activity (VAS), Patient/Parent Global Assessment of Overall Well-Being (VAS), JADAS 27-CRP, and JADAS 27-ESR are summarized in Table S3. Across these efficacy measures at week 12, improvement from baseline was observed in all age groups, and changes from baseline were consistent among the age groups. Changes in JADAS 27-ESR were similar to those for JADAS 27-CRP. Twenty-nine of 50 patients (58%) achieved JADAS 27-CRP ≤3.8, and 16 of 50 patients (32%) achieved JADAS 27-CRP ≤1 at week 12. The proportions of patients achieving these responses further increased at week 24 and were generally maintained through week 48 (Figure S2).

Safety. All AEs reported herein were treatment emergent. A summary of TEAEs through the cutoff date for this report is presented in Table 3. The median exposure time for the study drug (n = 57) was 412 days.

In part 1, eight patients (8 of 51, 15.7%) reported TEAEs, and no TEAEs were reported in more than one patient, except for headache, which was reported in two patients. No TEAEs were severe; the majority were reported as mild in severity. There were no TEAEs that led to study drug discontinuation, no AESIs, and no SAEs in part 1.

In parts 2 and 3, 52 of 57 patients (91.2%) reported TEAEs, with the highest proportion of reported TEAEs occurring in age group 12 to <18 years (19 of 19, 100%). The most frequently reported TEAEs (≥5 patients in any age group) were COVID-19 (23 of 57, 40.4%), upper respiratory tract infection (23 of 57, 40.4%), nasopharyngitis (13 of 57, 22.8%), pyrexia (10 of 57, 17.5%), abdominal pain (9 of 57, 15.8%), nausea (8 of 57, 14.0%), and arthralgia (5 of 57, 8.8%). The majority of TEAEs were mild or moderate in severity. Six patients (6 of 57, 10.5%) reported 12 SAEs, and two patients (2 of 57, 3.5%) reported TEAEs that led to study drug discontinuation; all were in the age group 12 to <18 years. The SAEs reported in more than one patient were abdominal pain (2 of 57, 3.5%) and nausea (2 of 57, 3.5%). SAEs considered by the investigator to have a reasonable possibility of being related to upadacitinib included upper respiratory tract infection, nausea, and conversion disorder. None of the SAEs led to discontinuation of study drug. The two nonserious TEAEs that led to study drug discontinuation included one event of JIA and one event of nasopharyngitis.

Reported AESIs were infrequent overall, with elevated creatine phosphokinase (CPK) and hepatic disorder being the most common (see Table 3). No AESIs were reported in part 1. As with overall TEAEs, rates of AESIs were higher in the older patient groups. Elevated CPK was asymptomatic in all patients. In part 2, serious and opportunistic infections, excluding tuberculosis and herpes zoster, included a single reported TEAE of upper respiratory tract infection in one patient that led to hospitalization and a single reported TEAE of esophageal candidiasis in one patient, respectively. Both were moderate in severity and resolved. Both events occurred on upadacitinib monotherapy, without concomitant use of other immunosuppressives. Single reported TEAEs of hepatic disorder included hepatosplenomegaly (with no elevated liver enzymes; moderate) in

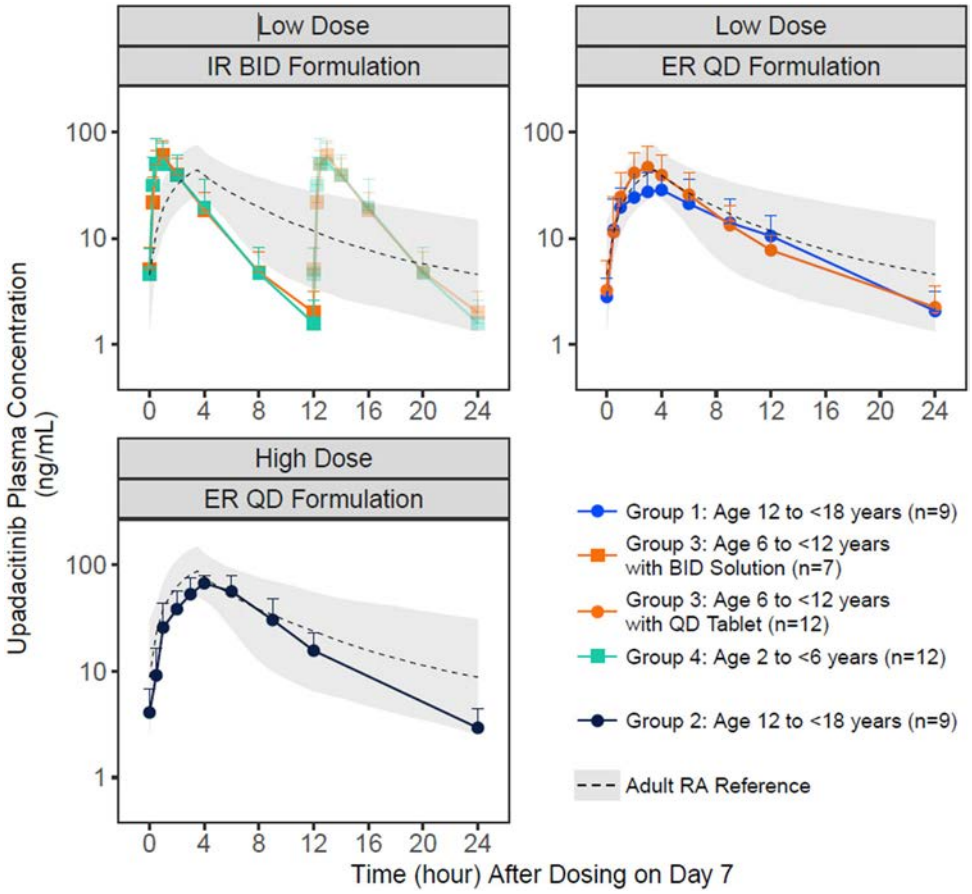


Figure 2. Mean + SD upadacitinib concentration-time profiles by group and dosing regimen (part 1). Adult RA reference represents median (dashed line) and (5th, 95th) percentiles (shaded area) of upadacitinib model-predicted plasma concentrations in adult patients with RA after administration of QD tablets. These reference pharmacokinetic profiles were simulated using a previously developed model for patients with RA.[12] The symbols with decreased opacity (over 12 to 24 hours) for the IR BID regimens are replicated from the observed data (over 0 to 12 hours) to account for difference in dosing frequency between the BID oral solution and QD tablet formulations. BID, twice-daily; ER, extended-release; IR, immediate-release; QD, once-daily; RA, rheumatoid arthritis; SD, standard deviation.

one patient and both alanine transaminase and aspartate aminotransferase increased in two patients (both mild for one patient and both severe for the other patient; see Table 3). All hepatic disorder events resolved, and none of the events were considered serious. The event of hepatosplenomegaly was attributed to patient's history of adipositas. AESIs of anemia, neutropenia, and lymphopenia were also reported; all were mild or moderate in severity. No patients had AESIs of malignancy, nonmelanoma skin cancer, gastrointestinal perforation, renal dysfunction, active tuberculosis, adjudicated major adverse cardiac cardiovascular events, or venous thromboembolisms during the study. There were no patients with a history of uveitis, and no uveitis events occurred during the study. No deaths were reported during the study.

DISCUSSION

In this open-label phase 1 study, upadacitinib administered as a fixed dose per body weight category was efficacious and well tolerated in pediatric patients with pcJIA. The observed upadacitinib pharmacokinetics in pediatric patients with pcJIA for the once-daily regimen using the ER formulation and the twice-daily regimen using the IR formulation were consistent with the characterized upadacitinib pharmacokinetics in adults for the respective formulations.[13,14] This study provided key information that supported approval of upadacitinib by the US Food and Drug Administration for the treatment of polyarticular JIA (pJIA) and pediatric PsA based on characterization of pharmacokinetics in pediatrics and extrapolation of upadacitinib efficacy from the respective adult populations.[15,16]

The original dosing regimen of upadacitinib was selected by leveraging pharmacokinetic data of upadacitinib in adult patients with RA and allometric scaling of clearance and volume of distribution based on body weight, with a goal of achieving upadacitinib exposures in patients with pcJIA similar to the exposures that were shown to be optimal in adult patients with RA. In this study, a preliminary population pharmacokinetic analysis indicated that the apparent oral clearance of upadacitinib in pediatric patients was underestimated when

Table 2. Pharmacokinetic parameters of upadacitinib by study group and dosing regimen, part 1*

Pharmacokinetic parameters (units)	Group 1: low dose, age 12–<18 y, QD tablet (n = 9)	Group 2: high dose, age 12–<18 y, QD tablet (n = 9)	Group 3: low dose, age 6–<12 y			Group 4: low dose, age 2–<6 y, BID solution (n = 12)[b]
			QD tablet (n = 12)	BID solution (n = 7)	Combined[a] (n = 19)	
C_{max} (ng/mL)	35.1 (37.2, 35)	69.8 (71.2, 19)	46.9 (52.3, 46)	58.9 (62.5, 35)	51.0 (56.1, 41)	46.6 (55.8, 56)
T_{max}[c] (h)	3.0 (1.0–6.0)	4.0 (2.0–6.0)	3.0 (1.0–6.0)[d]	1.0 (0.5–1.0)	–	1.0 (0.5–2.0)
Functional $t_{1/2}$[e] (h)	5.53 (1.77)[f]	4.79 (1.12)	5.20 (0.949)[d]	2.39 (0.277)	–	2.33 (0.351)[g]
AUC_{tau} (ng h/mL)	269 (282, 32)[f]	553 (572, 26)	318 (351, 41)[d]	198 (209, 36)	–	184 (217, 61)[g]
AUC_{0-24}[h] (ng h/mL)	269 (282, 32)[f]	553 (572, 26)	318 (351, 41)[d]	397 (417, 36)	346 (377, 38)[i]	369 (433, 61)[g]
CL_{ss}/F (L/h)	55.7 (58.2, 32)[f]	46.5 (49, 38)	41.6 (46.7, 67)[d]	19.7 (20.2, 25)	–	15.0 (16.8, 49)[g]
$CL_{ss}/F\_adjusted$[j] (L/h)	38.1 (39.8, 32)[f]	31.8 (33.5, 38)	28.5 (32.0, 67)[d]	19.7 (20.2, 25)	24.7 (27.4, 65)[i]	15.0 (16.8, 49)[g]

* Values are presented as geometric mean (arithmetic mean, % CV) unless otherwise specified. AE, adverse event; AUC, area under the plasma concentration-time curve from time 0 to 24 hours (AUC_{0-24}) or over a dose interval (AUC_{tau}); BID, twice-daily; C_{max}, maximal plasma concentration; CL_{ss}/F, apparent oral clearance at steady state; $CL_{ss}/F\_adjusted$, CL_{ss}/F adjusted for bioavailability; CV, coefficient of variation; ER, extended-release; functional $t_{1/2}$, functional half-life; IR, immediate-release; QD, once-daily; SD, standard deviation; T_{max}, time to maximal plasma concentration.
[a] C_{max}, AUC_{0-24}, and $CL_{ss}/F\_adjusted$ of the ER tablet and IR oral solution formulations are combined and summarized together for group 3.
[b] One patient was excluded for receiving an incorrect dose during the entire part 1 period, and another patient that did not receive a dose on day 7 due to an AE of vomiting was also excluded.
[c] Median (minimum through maximum).
[d] n = 11.
[e] Harmonic mean (pseudo-SD).
[f] n = 8.
[g] n = 10.
[h] For QD tablet, $AUC_{0-24} = AUC_{tau}$; for BID oral solution, $AUC_{0-24} = AUC_{tau} \times 2$.
[i] n = 18.
[j] CL_{ss}/F adjusted for the difference in bioavailability between formulations. For QD tablet, $CL_{ss}/F\_adjusted = 0.684 \times CL_{ss}/F$; for BID oral solution, $CL_{ss}/F\_adjusted = CL_{ss}/F$.

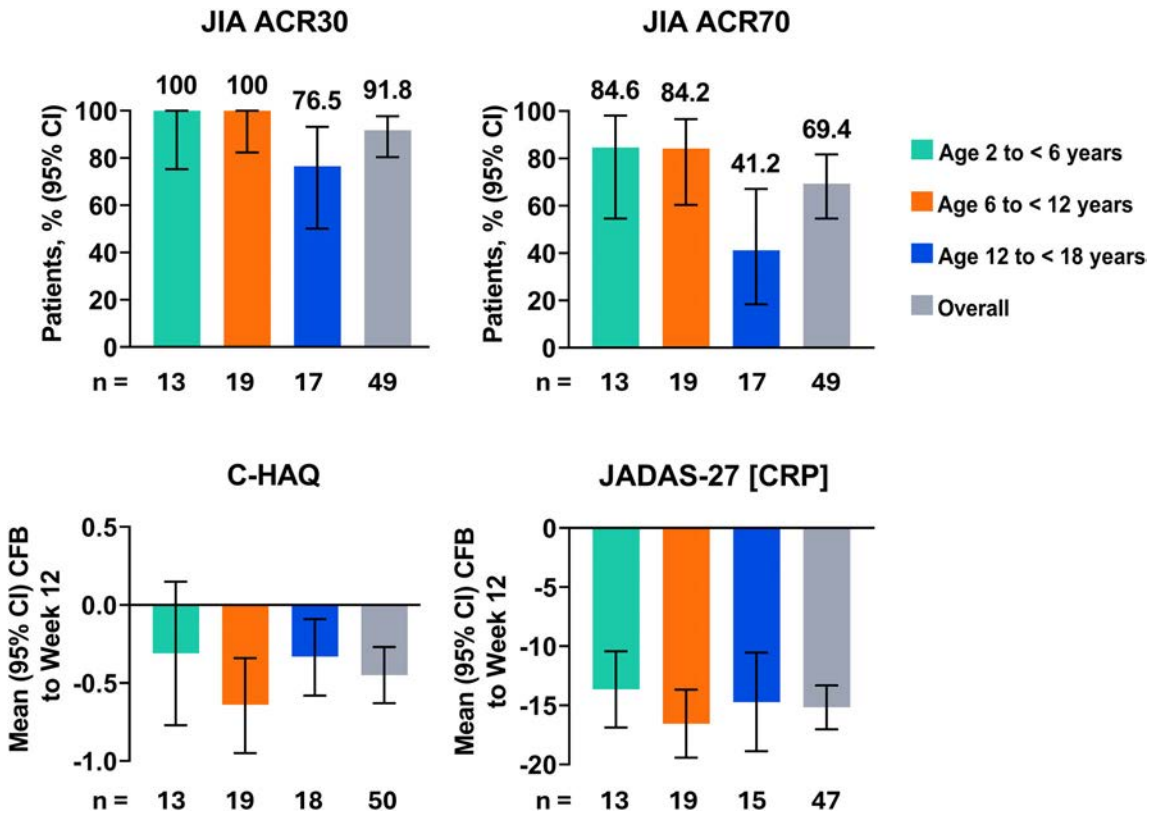


Figure 3. Interim analysis of efficacy data across all age groups (as observed) of upadacitinib at week 12 (parts 1 and 2). JIA ACR 30/70 indicate a 30% and 70% improvement from baseline, respectively, in three of six variables of JIA. Negative change from baseline indicates improvement in health or disease activity for C-HAQ and JADAS-27 [CRP]. ACR, American College of Rheumatology; C-HAQ, Childhood Health Assessment Questionnaire; CFB, change from baseline; JADAS-27 [CRP], Juvenile Arthritis Disease Activity Score based on C-reactive protein; JIA, juvenile idiopathic arthritis; 95% CI, 95% confidence interval.

Table 3. Safety interim analysis of upadacitinib up to 156 weeks for all patients treated in parts 1, 2, and 3*

TEAE	Patients with TEAE, n (%)			
	Age 2 to <6 years, n = 14	Age 6 to <12 years, n = 24	Age 12 to <18 years, n = 19	Total, n = 57
Any AE	12 (85.7)	22 (91.7)	19 (100)	52 (91.2)
Any serious AE	0	0	6 (31.6)	6 (10.5)
Treatment-related AE	2 (14.3)	8 (33.3)	11 (57.9)	21 (36.8)
AE leading to discontinuation of study drug	0	0	2 (10.5)	2 (3.5)
Deaths	0	0	0	0
AEs of special interest[a]				
Serious infections	0	0	1 (5.3)[b]	1 (1.8)[b]
Opportunistic infection	0	0	1 (5.3)[c]	1 (1.8)[c]
Hepatic disorder	0	2 (8.3)[d]	1 (5.3)[d]	3 (5.3)[d]
Anemia	1 (7.1)	0	0	1 (1.8)
Neutropenia	0	2 (8.3)	0	2 (3.5)
Lymphopenia	0	0	1 (5.3)	1 (1.8)
Elevated creatine phosphokinase	0	3 (12.5)	3 (15.8)	6 (10.5)

* A TEAE is defined as an adverse event with an onset date that is after the first dose of study drug, and no more than 30 days after the last dose of study drug. Opportunistic infections excluded tuberculosis, herpes zoster, anemia, and lymphopenia. Data cutoff date of September 22, 2022. AE, adverse event; MACE, major adverse cardiovascular event; TEAE, treatment-emergent adverse event.
[a] Other AEs of special interest included herpes zoster, active tuberculosis, malignancy (all types), adjudicated gastrointestinal perforations, renal dysfunction, adjudicated cardiovascular events (eg, MACE), and adjudicated embolic and thrombotic events (noncardiac, noncentral nervous system). None of these events were observed in the study.
[b] Upper respiratory tract infection.
[c] Candida esophagitis infection.
[d] One hepatosplenomegaly event, two high aspartate aminotransferase events, and two high alanine transaminase events. These events were transient, and patients fully recovered.

an estimate of upadacitinib clearance from adult patients with RA with a typical exponent of 0.75 was used to describe the relationship between body weight and clearance. Accordingly, a revised dosing regimen was developed and used in the younger pediatric patients (the majority of the 6 to <12 years and 2 to <6 years groups) to enable the attainment of upadacitinib exposures comparable with the target exposures in adults with RA. Based on previous analyses, the median (5th–95th percentile) upadacitinib AUC at steady state in adult patients with RA in phase 3 trials was 358 (234–701) and 708 (466–1,332) ng h/mL after daily administration of 15 mg and 30 mg ER tablets, respectively.[17] Compared with the target median adult exposure ($AUC_{0–24}$), the relative median upadacitinib $AUC_{0–24}$ within each group in this study ranged from ~0.77 to 1.07 (Table S5). In the younger pediatric patients, for whom the majority received revised doses, the observed upadacitinib exposures were nearly identical to the target exposures in adult patients with a ratio of 1.01.

In this study, an IR twice-daily oral solution formulation was used for patients with lower body weight (ie, <30 kg) or unable to swallow tablets. Despite distinct pharmacokinetic profiles due to different dosing frequency, the twice-daily oral solution is expected to provide similar efficacy in pcJIA relative to the once-daily tablet if similar upadacitinib $AUC_{0–24}$ can be achieved. This is supported by previously conducted exposure-response analyses of key efficacy end points in adult patients with RA, for whom a model developed based on data from a twice-daily capsule formulation in phase 2 studies successfully predicted the observed efficacy of the once-daily tablet formulation in phase 3 studies.[18] The analyses showed that the upadacitinib twice-daily IR and once-daily ER regimens providing similar $AUC_{0–24}$ were predicted to achieve similar efficacy responses. Notably, the limited impact of short-term drug concentration fluctuation on clinical outcome in RA has also been demonstrated in another JAK inhibitor, tofacitinib, where exposure-response analyses of clinical and nonclinical data suggested that daily AUC is more correlated with clinical response[19] than minimum or maximum daily concentration.

Based on the descriptive analyses of efficacy end points, improvements in measures of pcJIA disease activity, pain, function, and overall well-being were observed after administration of upadacitinib at week 12 in this study and were generally maintained through week 48. Improvements with upadacitinib were observed across the evaluated age groups, which included patients with pcJIA ages 2 to <18 years old, with numerically higher response rates in patients ages 2 to <12 years than those 12 to <18 years. Notably, most patients in the oldest age group had longer disease duration and more prior bDMARD exposure, which may be associated with numerically lower response rates (Table S4). In comparison, most patients in the younger age groups had shorter disease duration and less or no prior bDMARD exposure (Table 1).

There were no patients with a history of uveitis, and no uveitis events occurred during the study. The safety profile of upadacitinib in pediatric patients with pcJIA was generally consistent with the currently known safety profile of upadacitinib in adults and adolescents with inflammatory conditions.

The descriptive summary of efficacy of upadacitinib in patients with pcJIA can be contextualized with results reported by other JAK inhibitor trials. Tofacitinib and baricitinib have demonstrated efficacy for treatment of pcJIA in phase 3 studies with double-masked withdrawal designs.[8,9] Before the double-masked period of these phase 3 studies, there was a run-in period during which patients received study drugs in an open-label manner, which was similar to our open-label study. At the end of the run-in period, tofacitinib and baricitinib achieved JIA ACR 30, 50, and 70 responses of 77%, 70%, and 51% at week 18 and 76%, 64%, and 46% at week 12, respectively. In our study, numerically higher responses were observed, although the sample size for this analysis was relatively small. Of note, prior exposure to conventional synthetic DMARDs and bDMARDs may differ and limit comparison of the results across studies.

This study has several limitations. The primary objective of this open-label, nonrandomized phase 1 study was the assessment of pharmacokinetics and safety as well as descriptive analyses of efficacy. Thus, it was not set up to generate direct comparative information against other active therapies for pcJIA or placebo. Additionally, the number of patients enrolled is relatively small, which may not allow robust conclusions on safety. However, the results from this study, along with the extensive safety data available for upadacitinib, demonstrated consistency of the upadacitinib safety profile between pediatric patients and adults and encouraged the recent regulatory approvals of upadacitinib in the treatment of pJIA and pediatric PsA.[15] Of note, a long-term safety registry study, as well as longer-term data from the current study, will help further inform upadacitinib safety in pcJIA. Another limitation of this study is that there was a dose revision during the study that may potentially confound the study outcome. The decision for dose revision was made based on a preplanned analysis of available pharmacokinetic data collected from this study and another phase 1 study in pediatric patients with atopic dermatitis. The dose revision was implemented to ensure the attainment of sufficient upadacitinib exposure in pediatric patients, which was reflected by the higher observed exposures in the two younger groups (Figure 2).

Upadacitinib was well tolerated and efficacious in patients with pcJIA across all age groups (age 2 to <18 years) with plasma exposures comparable to adult patients with RA at the evaluated dosing regimens. No new safety risks were observed in patients with pcJIA, and the benefit-risk profile of upadacitinib was assessed as favorable based on the safety and efficacy outcomes of the study to date.

ACKNOWLEDGMENT

The authors thank Stormy Koeniger, PhD, an employee with AbbVie, for medical writing support.

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 Mohamed 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

AbbVie participated in the study design; in the data collection, analysis, and interpretation; in the writing of the manuscript; and in the decision to submit the manuscript for publication. Publication of this article was contingent upon approval by AbbVie. No honoraria or payments were made for authorship.

REFERENCES

1. Petty RE, Southwood TR, Manners P, et al. International league of associations for rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31(2):390–392.
2. Barut K, Adrovic A, Sahin S, et al. Juvenile idiopathic arthritis. Balkan Med J 2017;34(2):90–101.
3. Martini A, Lovell DJ, Albani S, et al. Juvenile idiopathic arthritis. Nat Rev Dis Primers 2022;8(1):5.
4. Macaubas C, Nguyen K, Milojevic D, et al. Oligoarticular and polyarticular JIA: epidemiology and pathogenesis. Nat Rev Rheumatol 2009;5(11):616–626.
5. Brunner HI, Schanberg LE, Kimura Y, et al; PRCSG Advisory Council and the CARRA Registry Investigators. New medications are needed for children with juvenile idiopathic arthritis. Arthritis Rheumatol 2020;72(11):1945–1951.
6. Glerup M, Rypdal V, Arnstad ED, et al; Nordic Study Group of Pediatric Rheumatology. Long-term outcomes in juvenile idiopathic arthritis: eighteen years of follow-up in the population-based Nordic juvenile idiopathic arthritis cohort. Arthritis Care Res (Hoboken) 2020;72(4):507–516.
7. Rinvoq (upadacitinib). Package insert. AbbVie, Inc.; 2019. Revised 2024.
8. Ruperto N, Brunner HI, Synoverska O, et al; Paediatric Rheumatology International Trials Organisation (PRINTO) and Pediatric Rheumatology Collaborative Study Group (PRCSG). Tofacitinib in juvenile idiopathic arthritis: a double-blind, placebo-controlled, withdrawal phase 3 randomised trial. Lancet 2021;398(10315):1984–1996.
9. Ramanan A, Quartier P, Okamoto N, et al. LB0002 Baricitinb in juvenile idiopathic arthritis: a phase 3, double-blind, placebo-controlled, withdrawal, efficacy and safety study. Ann Rheum Dis 2022;81(Suppl 1):207–208.
10. Ramanan AV, Quartier P, Okamoto N, et al; JUVE-BASIS investigators; Paediatric Rheumatology International Trials Organisation. Baricitinib in juvenile idiopathic arthritis: an international, phase

3, randomised, double-blind, placebo-controlled, withdrawal, efficacy, and safety trial. Lancet 2023;402(10401):555–570.

11. Giannini EH, Ruperto N, Ravelli A, et al. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40(7):1202–1209.
12. Klünder B, Mittapalli RK, Mohamed MF, et al. Population pharmacokinetics of upadacitinib using the immediate-release and extended-release formulations in healthy subjects and subjects with rheumatoid arthritis: analyses of phase I-III clinical trials. Clin Pharmacokinet 2019;58(8):1045–1058.
13. Mohamed MF, Camp HS, Jiang P, et al. Pharmacokinetics, safety and tolerability of ABT-494, a novel selective JAK 1 inhibitor, in healthy volunteers and subjects with rheumatoid arthritis. Clin Pharmacokinet 2016;55(12):1547–1558.
14. Mohamed MF, Zeng J, Marroum PJ, et al. Pharmacokinetics of upadacitinib with the clinical regimens of the extended-release formulation utilized in rheumatoid arthritis phase 3 trials. Clin Pharmacol Drug Dev 2019;8(2):208–216.
15. NDA 211675/S-021 and S-022. US Food and Drug Administration. April 26, 2024. Accessed September 18, 2024. https://www.accessdata.fda.gov/drugsatfda_docs/appletter/2024/211675Orig1s021,%20s022ltr.pdf
16. Qian Y, Schlachter L, Eckert D, et al. Extrapolation of upadacitinib efficacy in juvenile idiopathic arthritis leveraging pharmacokinetics, exposure–response models, and real-world patient data. Clin Pharmacol Ther 2024;116(6):1635–1645.
17. Muensterman E, Engelhardt B, Gopalakrishnan S, et al. Upadacitinib pharmacokinetics and exposure-response analyses of efficacy and safety in psoriatic arthritis patients - analyses of phase III clinical trials. Clin Transl Sci 2022;15(1):267–278.
18. Mohamed MF, Klünder B, Camp HS, et al. Exposure-response analyses of upadacitinib efficacy in phase II trials in rheumatoid arthritis and basis for phase III dose selection. Clin Pharmacol Ther 2019;106(6):1319–1327.
19. Lamba M, Hutmacher MM, Furst DE, et al. Model-informed development and registration of a once-daily regimen of extended-release tofacitinib. Clin Pharmacol Ther 2017;101(6):745–753.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 594–603
DOI 10.1002/acr.25476

REVIEW ARTICLE

Relationship Between Number of Different Lower-Limb Resistance Exercises Prescribed in a Program and Exercise Outcomes in People With Knee Osteoarthritis: A Systematic Review With Meta-Regression

Belinda J. Lawford,[1] Kim L. Bennell,[1] Libby Spiers,[1] Alexander J. Kimp,[1] Andrea Dell'Isola,[2] Alison R. Harmer,[3] Martin Van der Esch,[4] Michelle Hall,[3] and Rana S. Hinman[1]

Objective. We determine whether there is a relationship between the number of different lower-limb resistance exercises prescribed in a program and outcomes for people with knee osteoarthritis.

Methods. We used a systematic review with meta-regression. We searched the Cochrane Central Register of Controlled Trials, MEDLINE, and Embase up to January 4, 2024. We included randomized controlled trials that evaluated land-based resistance exercise for knee osteoarthritis compared with nonexercise interventions. We conducted meta-regressions between number of different exercises prescribed and standardized mean differences (SMDs) for pain and function. Covariates (intervention duration, frequency per week, use of resistance exercise machine[s], and comparator type) were applied to attempt to reduce between-study heterogeneity.

Results. Forty-four trials (3,364 participants) were included. The number of resistance exercises ranged from 1 to 12 (mean ± SD 5.0 ± 3.0). Meta-regression showed no relationship between the number of prescribed exercises and change in pain (slope coefficient: −0.04 SMD units [95% confidence interval {95% CI} −0.14 to 0.05]) or self-reported function (SMD −0.04 [95% CI −0.12 to 0.05]). There was substantial heterogeneity and evidence of publication bias. However, even after removing 31 trials that had overall unclear/high risk of bias, there was no change in relationships.

Conclusion. There was no relationship between the number of different lower-limb resistance exercises prescribed in a program and change in knee pain or self-reported function. However, given that we were unable to account for all differences in program intensity, progression, and adherence, as well as the heterogeneity and overall low quality of included studies, our results should be interpreted with caution.

INTRODUCTION

Knee osteoarthritis (OA) affects >654 million people aged ≥40 years worldwide.[1] All current clinical guidelines advocate exercise for management of knee OA, irrespective of age, comorbidity, pain severity, or disability.[2–6] Although numerous systematic reviews support the effectiveness of exercise for knee OA, effect sizes are small to moderate and decline over time.[7,8] New ways of enhancing the effectiveness of exercise are needed, such as by identifying the optimal content and dosage of exercise programs.

Muscle weakness is common in people with OA, often evident in muscles surrounding the affected joint,[9–11] and is associated with pain and physical dysfunction.[12] Improving muscle strength via resistance training is therefore a common focus of exercise for knee OA management and is hypothesized to be one of the mechanisms by which exercise leads to improvements in symptoms.[13–16] The American College of Sports Medicine (ACSM) provides specific recommendations for prescription of resistance exercise programs for people with arthritis, including the frequency, intensity, and duration of the program.[17] However,

Drs Bennell and Hinman's work was supported by the National Health and Medical Research Council (grants 1174431 and 2025733).

[1]Belinda J. Lawford, PhD, Kim L. Bennell, PhD, Libby Spiers, BPhysio, Alexander J. Kimp, DPT, Rana S. Hinman, PhD: The University of Melbourne, Melbourne, Victoria, Australia; [2]Andrea Dell'Isola, PhD: Lund University, Lund, Sweden; [3]Alison R. Harmer, PhD, Michelle Hall, PhD: The University of Sydney, Sydney, New South Wales, Australia; [4]Martin Van der Esch, PhD: Amsterdam University of Applied Sciences, Amsterdam, The Netherlands.

Drs Hall and Hinman 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.25476).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25476.

Address correspondence via email to Belinda J. Lawford, PhD, at belinda.lawford@unimelb.edu.au.

Submitted for publication July 1, 2024; accepted in revised form November 22, 2024.

SIGNIFICANCE & INNOVATIONS

- Although numerous systematic reviews support the effectiveness of exercise for knee osteoarthritis (OA), effect sizes are small to moderate and decline over time. New ways of enhancing the effectiveness of exercise are needed, such as by identifying the optimal content and dosage of exercise programs. One element of resistance exercise dosage that has not yet been evaluated before in OA is the number of different prescribed lower-limb resistance exercises.
- We found that there was no evidence of a relationship between the number of prescribed exercises and change in pain or self-reported function.
- However, there was substantial heterogeneity and evidence of publication bias, and many included trials were of low quality. As such, our results should be interpreted with caution.

systematic reviews and randomized controlled trials (RCTs) have found that the effects of exercise on OA symptoms are similar across different dosages,[18–23] and thus, no optimal dosage has yet been established. Recently, the 2023 EULAR recommendations for management of hip and knee OA highlighted optimal dosage of exercise as a research priority.[6]

One element of resistance exercise dosage that has not yet been evaluated in an RCT or systematic review in OA is the number of different prescribed lower-limb resistance exercises. The ACSM guidelines recommend that resistance exercise programs target all major muscle groups,[17] although no specific guidance is provided as to the number of different exercises that should be included. It is possible that a program that includes multiple exercises, targeting multiple different muscle groups, may lead to greater improvements in symptoms than a program that comprises fewer exercises. However, it is also plausible that a program with fewer different resistance exercises may be less burdensome, be perceived by people as being easier to undertake,[24–27] and potentially promote better long-term engagement and therefore greater benefits on symptoms. Indeed, research outside of OA has suggested that brief, simple 'minimal-dose' resistance exercise programs are effective at increasing muscle mass and improving function.[24,28] Thus, the objective of this systematic review was to evaluate whether there is a relationship between the number of different lower-limb resistance exercises prescribed to people with knee OA and improvements in pain and self-reported function.

METHODS

This review used data extracted as part of our updated Cochrane Systematic Review evaluating the effectiveness of land-based exercise for knee OA.[29] Our approach follows the Cochrane Handbook for Systematic Reviews of Interventions.[30] This manuscript is written in accordance with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement.[31] Data are available upon reasonable request to corresponding author.

Literature search. We searched three databases (MEDLINE OvidSP, Embase OvidSP, and the Cochrane Central Register of Controlled Trials [CENTRAL]) from inception to January 4, 2024. No language restrictions were applied. The search strategy is shown in Supplementary Material S1. Gray literature was not searched.

Eligibility criteria. We included RCTs involving adults with an established diagnosis of knee OA according to accepted clinical criteria[4,32,33] or who self-reported knee OA on the basis of chronic joint pain (with or without radiographic confirmation). We also included trials that included participants with OA in other joints, provided that the outcomes for people with knee OA were presented separately (or could be obtained from the trial authors), or if >80% of participants in the trial had knee OA. Eligible RCTs were required to include assessment of self-reported knee pain severity and/or function.

We included any land-based resistance exercise regimens. Eligible interventions were required to clearly report the number of different resistance exercises prescribed in the program. Exercise interventions that combined resistance exercise with any other type of exercise (eg, aerobic, balance) were not eligible. The only nonresistance exercise permitted in an eligible intervention was in a warm-up or warm-down. Eligible exercise interventions could incorporate education and/or behavior change strategies designed to help participants adhere to the exercise program (eg, information about benefits of exercise, short-message services or mobile app to maximize adherence), could be supervised or unsupervised, and could include other nonexercise co-interventions alongside the exercise intervention, as long as these co-interventions were permitted/provided similarly in any comparator groups. Exercise interventions were ineligible if they were perioperative (ie, specifically recruited all participants from surgical waiting lists and/or evaluated outcomes postsurgery), included whole-body vibration, or included any other type of exercise that was not resistance-based strengthening (eg, aerobic, balance, neuromuscular).

Comparator groups eligible for inclusion were the following:

- Placebo, sham, or attention control (any intervention that was designed to control for contextual/placebo effects and described by the authors as 'placebo' or 'sham,' and/or an 'attention control' involving more than one bout of synchronous interaction with study personnel/care provider [excluding contact to obtain outcome measures]).

- No treatment (no defined allocated intervention), usual care (stated that participants could receive normal care, but not controlled by the trial), or minimal education/information (provided with a one-off information resource).
- Any nonexercise nonsurgical intervention that was also offered/provided equally as a co-intervention in the exercise group (eg, weight loss, manual therapy, insoles/footwear, pain coping skills, physical therapy that does not include an exercise component).

Study selection. Teams of two review authors independently screened titles and abstracts for inclusion. We retrieved the full-text study reports/publication of any study that either of the two reviewers considered potentially eligible. Two review authors independently screened the full text and identified studies for inclusion. If agreement was not achieved at any stage, a third review author (BJL or MH) adjudicated.

Quality assessment. The Cochrane Collaboration's Risk of Bias 1 tool for assessing risk of bias in RCTs was used to assess potential bias.[34] Two authors (BJL and MH) independently assessed risk of bias for each included study. Disagreements were resolved through discussion or by involving another review author. The following risk of bias domains were assessed as adequate (low risk of bias), inadequate (high risk of bias), or unclear (insufficient information): random sequence generation (selection bias), allocation concealment (selection bias), masking of participants and personnel (performance bias), masking of outcome assessment (detection bias), incomplete outcome data (attrition bias), and selective outcome reporting (reporting bias).

Data collection. Teams of two review authors independently extracted outcome data from included studies using a data collection form in Covidence, which was pilot tested on at least two eligible studies. Disagreements were resolved by consensus or by involving a third review author. We extracted data relating to study participants (number randomized to each group, mean age, percentage who were female, diagnostic criteria, and body mass index [BMI]), comparison and exercise interventions, and outcomes of pain and self-reported function.

We extracted the total number of different lower-limb resistance exercises prescribed in a single-exercise program. If a range of different resistance exercises were prescribed either within a single session or across the intervention, the average was calculated and used for meta-analysis and meta-regression.

For outcomes of pain and self-reported function, we extracted means and SDs. When authors used more than one pain or self-reported function outcome, we used the hierarchy of outcomes used in the prior Cochrane review of exercise for knee OA.[7] Outcomes were assessed immediately at the end of the treatment (post-treatment). When trialists reported both end score and change from baseline values for the same outcome, we extracted the end score values. We extracted intention-to-treat (ITT)–analyzed data, if reported. We contacted investigators as needed to obtain missing outcome data (eg, when data were not available for all participants). If no response from investigators was received after two attempts (at least one month apart), data were considered irretrievable. Where possible, we calculated missing SDs from other statistics, such as standard errors, confidence intervals, or *P* values, according to the methods recommended in the Cochrane Handbook for Systematic Reviews.[35]

Data analysis. We used the Cochrane Collaboration statistical software, Review Manager,[36] to calculate standardized mean differences (SMDs) with 95% confidence intervals (95% CIs). We entered data with a consistent direction of effect across studies and calculated the SMD based on the number of patients randomized at baseline. When multiple arms were reported in a single trial, we included only the relevant arms. If a three-arm trial was included (eg, resistance exercise A vs resistance exercise B vs no intervention), the resistance exercise groups for that trial were pooled to calculate the SMD.

We performed meta-regressions in SPSS (version 29, IBM Statistics) to evaluate the relationship between number of different resistance exercises prescribed and SMDs for pain and self-reported function. Study heterogeneity was assessed using the I^2 statistic. Because we anticipated heterogeneity in SMDs due to differences in study characteristics, random-effects analyses were used. We included a number of study-level covariates (chosen based on theoretical plausibility and on precedents for other exercise meta-regressions[37]) in an attempt to reduce between-study heterogeneity. These included (1) intervention duration, (2) exercise frequency per week, (3) use (or not) of resistance exercise machine(s), and (4) type of comparator (attention control/placebo, no treatment/usual care/limited education, or a co-intervention that is also equally applied in the exercise group). Other dosage variables (eg, intensity, number of sets and repetitions), as well as information related to program adherence, were considered but were poorly and inconsistently reported across included trials (Supplementary Table S1) and as such were not included as covariates in the main analysis. However, we conducted sensitivity analyses to determine whether meta-regressions differed when including a variable relating to exercise volume for trials where that data were available (number of sets × number of repetitions × frequency per week). We also ran meta-regressions without any covariates and conducted additional sensitivity analyses to determine whether results from meta-regressions differed when low-quality trials (those deemed to be at unclear or high risk of bias, defined as having unclear/high risk of bias on three or more of six bias domains) were excluded. To examine publication bias, we conducted Egger's regression test[38] and inspected funnel plots.

For completeness, we also conducted secondary meta-analyses in which we subgrouped all studies by number of

prescribed resistance exercises, allowing us to determine the SMD in pain and self-reported function for each number of prescribed exercises. We also conducted secondary meta-regressions to examine whether effects differed based on program supervision (entirely supervised vs mix of supervised and unsupervised) and use of resistance machines (included any machine-based exercises vs no machine-based exercises, where 'machine' was defined as a fixed resistance exercise machine that targeted specific muscle groups).

RESULTS

Study selection. The literature search resulted in 10,314 articles, with 10,254 being screened on title and abstract after removal of duplicates. After screening of 759 records in full text, 44 trials, with 3,206 participants, were included in the final analyses (Figure 1).

Study characteristics. Supplementary Table S1 provides an overview of the study characteristics. Participants were, on average, 61.5 years old (SD 6.1; range, 49.9–71.8 years), with a mean ± SD BMI of 29.4 ± 2.7 kg/m^2 (range, 23.3–34.3). The proportion of women ranged from 40% to 100% (mean ± SD 72% ± 17%). The mean ± SD intervention duration was 11.5 ± 11.1 weeks (range, 4–78 weeks). Fifteen trials (34%) used resistance exercise machines.[39–52] Nine trials (20%) compared exercise with attention control/placebo.[41,45,46,50,53–57] Twenty trials (45%) compared exercise with usual care, no intervention, or limited education.[39,40,42–44,48,49,52,58–69] Fifteen trials (34%) compared exercise with a co-intervention (that was also applied equally in the exercise group).[47,51,70–82] The mean ± SD number of prescribed lower-limb resistance exercises was 5.0 ± 3.0 (range 1–12).

All but three trials (7%) reported pain.[42,75,79] Twenty-four studies (55%) measured pain using the Knee Injury and

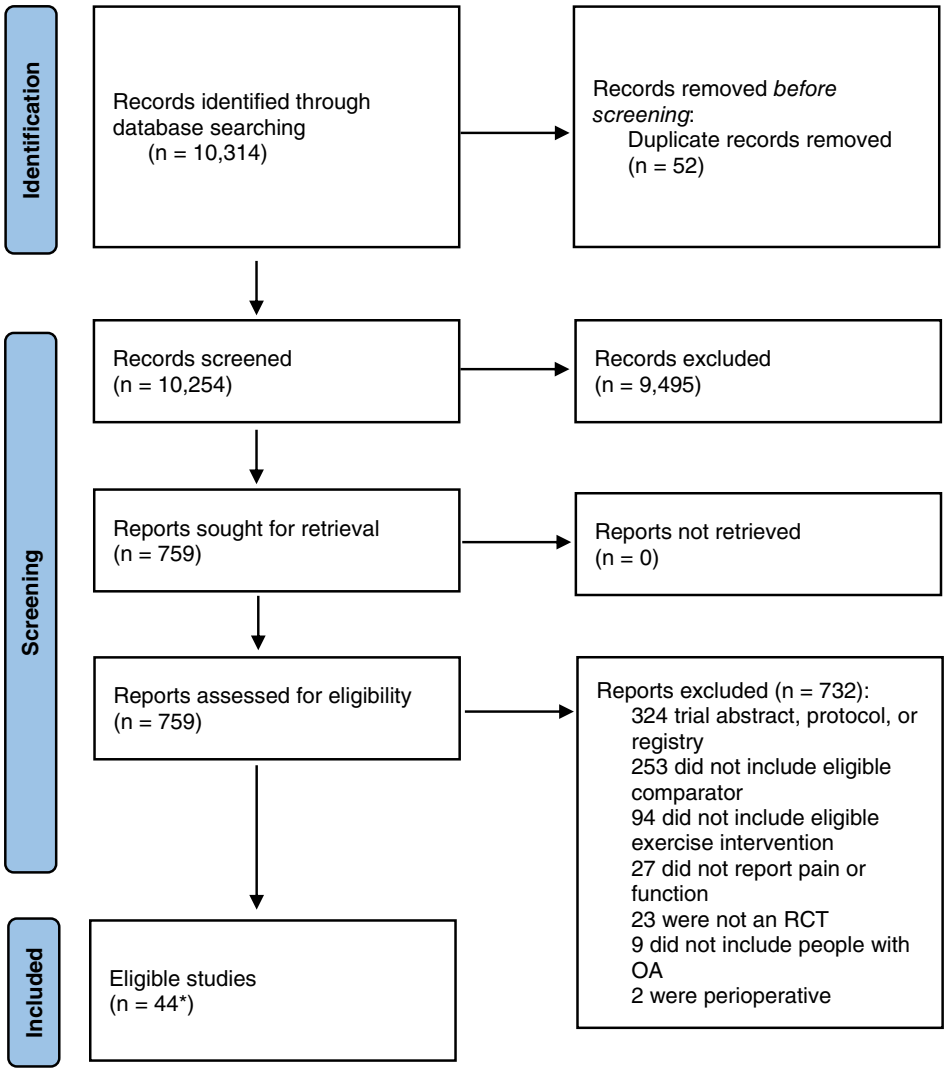


Figure 1. Flowchart of RCTs included in the systematic review and meta-analysis. * Including an additional 17 trials from the prior version of the Cochrane review. OA, osteoarthritis; RCT, randomized controlled trial.

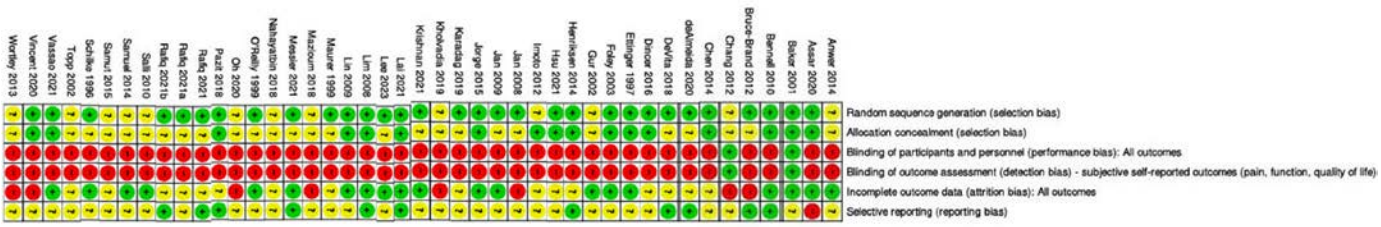


Figure 2. Risk of bias of included trials. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25476/abstract.

Osteoarthritis Outcome Score (KOOS) or Western Ontario and McMaster Universities Arthritis Index (WOMAC).[40,41,44–46,51–53,59,61,63,65,67–69,72,74,76–80,82,83] Fifteen studies (34%) measured pain using a visual analog scale (VAS) or numeric analog scale (NRS).[39,43,47,48,50,54,56–58,60,62,64,70,71,73] One study (2%) used a Likert scale,[55] one (2%) used the Lequesne index,[66] and one (2%) used the Osteoarthritis Screening Index (OASI).[49]

All but five trials (11%) reported function.[50,57,64,66,71] Thirty-two studies (73%) measured self-reported function using the KOOS Activities of Daily Living or WOMAC function subscale.[40–48,51–54,58,59,61–63,65,68–70,72–79,81,82] One study (2%) used a 1 to 6 scale,[55] one (2%) used VAS/NRS,[60] one (2%) used the Lequesne index,[39] one (2%) used the OASI,[49] one (2%) used a patient-specific functional scale,[80] and one (2%) used the Short Form Survey (SF-36).[56]

Risk of bias. Figure 2 displays the risk of bias across included trials. Sixteen studies (36%) adequately generated a random sequence and concealed the sequence until allocation; thus, we considered these studies at low risk of selection bias.[39,41,43,44,50,52,53,55,57,58,61,65,68,71,73,74] Two studies (5%) adequately masked participants and investigators to their treatment by using limited disclosure and/or an adequate comparison intervention that enabled masking.[53,72] Thus, we considered these studies at low risk of performance bias. Attrition bias was not likely in 19 studies (43%) because the number of dropouts was small (<20%) and consistent across interventions and/or authors used ITT analyses.[41–44,46,47,53,55,57,58,60,63–65,68,70,71,77,82] Reporting bias was unlikely in 11 studies (25%) because these were prospectively registered and results for all registered outcomes were reported.[40,46,52,54,57–59,61,65,80,81]

Relationship between number of different prescribed resistance exercises and pain. Meta-regression showed there was no evidence of a relationship between number of prescribed lower-limb resistance exercises and change in pain (slope coefficient: –0.04 SMD units [95% CI –0.14 to 0.05]; P = 0.36; Figure 3). However, there was substantial heterogeneity (I^2 = 97%). After removing 31 trials (70%) with unclear or high overall risk of bias,[39,40,42,45,47–51,54,56,59,60,62–64,66,67,70,72–82] there was still no evidence of a relationship (SMD –0.01 [95% CI –0.07 to 0.06]; P = 0.82; I^2 = 12%; Supplementary Figure S1). Supplementary Table S2 and Supplementary Figure S2 report meta-analysis results, which showed that 95% CIs of SMD estimates overlapped for each number of prescribed resistance exercises. Egger's test indicated there was risk of publication bias (Supplementary Figure S3). There were no differences in results of meta-regressions when not including any covariates (Supplementary Material S2) or when including a covariate for exercise volume (Supplementary Material S3). There was no evidence of an association between number of exercise and outcomes according to whether the program was entirely supervised or included a mix of supervised and unsupervised sessions or whether resistance exercises were machine-based or not (Supplementary Figures S4–S7).

Relationship between number of different prescribed resistance exercises and self-reported function. Meta-regression showed no evidence of a relationship between number of prescribed lower-limb resistance exercises and change in self-reported function (SMD –0.04 [95% CI –0.12 to 0.05]; P = 0.40; Figure 3). However, there was substantial heterogeneity (I^2 = 82%). After removing 31 trials (70%) with unclear or high overall risk of bias,[39,40,42,45,47–51,54,56,59,60,62–64,66,67,69,70,72–82] there was still no evidence of a relationship (SMD 0.09 [95% CI –0.07 to 0.25]; P = 0.20; I^2 = 0%; Supplementary Figure S1). Supplementary Table S2 and Supplementary Figure S2 report meta-analysis results, which showed that confidence intervals of SMD estimates overlapped for each number of prescribed resistance exercises. Egger's test indicated there was no risk of publication bias (Supplementary Figure S3). There were no differences in results of meta-regressions when not including any covariates (Supplementary Material S2) or when including a covariate for exercise volume (Supplementary Material S3). There was no evidence of an association between number of exercise and outcomes according to whether the program was entirely supervised or included a mix of supervised and unsupervised sessions or whether resistance exercises were machine-based or not (Supplementary Figures S4–S7).

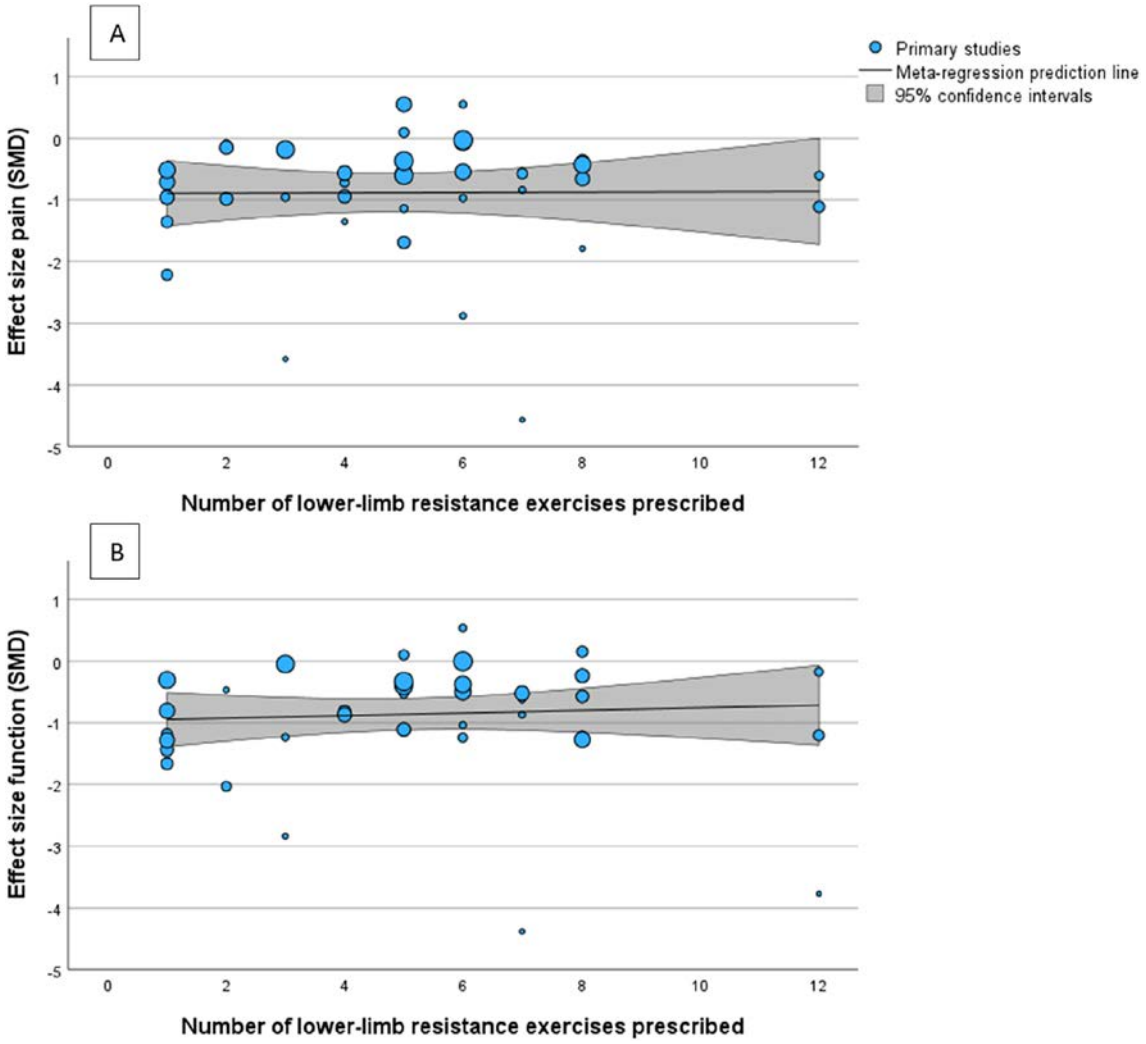


Figure 3. Meta-regression analysis: SMD of included studies according to pain (A) and self-reported function (B) at increasing numbers of prescribed lower-limb resistance exercises. Lower SMDs favor exercise. Covariates included in analysis were (1) intervention duration, (2) exercise frequency per week, (3) use (or not) of resistance exercise machine(s), and (4) type of comparator (attention control/placebo, no treatment/usual care/limited education, or a co-intervention that is also equally applied in the exercise group). Forty trials included pain (I^2 = 96%). 39 trials included self-reported function (I^2 = 77%). SMD, standardized mean difference. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25476/abstract.

DISCUSSION

The objective of this systematic review was to evaluate whether there is a relationship between the number of different land-based lower-limb resistance exercises prescribed to people with knee OA and clinical outcomes. We found no evidence of a relationship between the number of prescribed exercises and change in knee pain or self-reported function. Exercise programs with fewer different resistance exercises appeared to be just as effective as programs with numerous different exercises. However, we were unable to account for exercise intensity, progression or adherence to exercise prescriptions due to poor reporting, and the overall quality of included trials was considered low (mostly unclear or high risk of bias). Thus, we have low certainty in our results, and our findings should be interpreted with caution.

To our knowledge, this is the first systematic review evaluating the relationship between the number of different resistance exercises in a prescribed program and change in pain and self-reported function in OA. Our findings suggest that the number of different resistance exercises prescribed may not be an important element of effective exercise programs for OA symptoms. This is indirectly supported by RCTs[19–21,46] and meta-analyses[18,22,23,29,37,84] that found no relationship between other exercise dosage variables, including exercise intensity and frequency per week, and clinical outcomes from exercise. Collectively, this suggests that an exercise dose response may not exist in people with knee OA with respect to clinical outcomes of pain and self-reported function. However, it is possible that a greater number of resistance exercises may have more benefit on other outcomes that we did not examine, such as muscle strength or objective

functional performance measures. Recent systematic reviews have shown that 'low-dose' resistance programs (comprising only 1–5 different exercises) are effective at increasing muscle strength and other fitness outcomes.[85] Recent work examining mechanisms of effects of exercise showed that improvements in muscle strength only mediate a small proportion of the effects on changes in pain and function,[16] and psychosocial factors, including self-efficacy, pain beliefs, and fear of movement, may play a more important role.[86,87] There is also increasing evidence that clinical improvements associated with exercise may be largely driven by contextual factors and placebo response,[88–90] rather than specific effects related to exercise itself.

Given that we found no relationship between number of different resistance exercise prescribed and clinical outcomes, simple resistance programs with fewer exercises may have some advantages over programs that involve a greater number of exercises. One such advantage may be better exercise adherence, although we were unable to explore whether programs with fewer prescribed exercises had better adherence relative to programs with more exercises due to inconsistent reporting of adherence among included trials (Supplementary Table S1). Research in people with neck and low back pain has previously found that a lower number of prescribed exercises was associated with better long-term adherence.[91,92] Programs that involve fewer different exercises may also keep perceived burden to a minimum,[93] overcoming time-related barriers to exercise[94] and ensuring that quality of exercise performance is maintained.[95] Our findings suggest that health care providers should prescribe their patients with knee OA a program that meets their individual preferences, including the number of different exercises that the person is willing and able to accommodate into their daily routines. However, we acknowledge that our findings should be interpreted with caution given that we were unable to account for exercise intensity and progression over time. Prior research has indicated that physiotherapists are skeptical about the effectiveness of 'simplified' exercise programs involving a single resistance exercise,[96] viewing single-exercise programs as being rigid and limiting use of their professional skills. In addition, completing a range of different resistance exercises may have important benefits for other domains of health, such as sarcopenia and physical frailty, where a combination of upper- and lower-limb exercises that target different major muscle groups are recommended to increase strength and muscle mass.[97,98]

We have low certainty in our findings. Although we included numerous study-level covariates in our meta-analyses in an attempt to reduce between-study heterogeneity, I^2 values were still large, indicating substantial heterogeneity. We also found evidence of publication bias for pain outcomes, suggesting potential overestimation of treatment effects. Finally, most included trials were of poor quality (70% were deemed to be at overall unclear/high risk of bias), and 43% used small sample sizes (<50 participants). Future exercise trials should use robust designs to reduce risk of bias, such as concealing randomization schedules, analyzing data from all randomized participants (regardless of adherence to the intervention), and prospectively registering the trial in a clinical trials registry. Improved reporting of exercise interventions is also required because important intervention details were often inadequately described, including information relating to exercise dosage, delivery, and adherence (Supplementary Table S1). Trial authors should follow the Consolidated Standards of Reporting Trials checklist when presenting results[99] and use the Consensus on Exercise Reporting Template[100] and template for intervention description and replication[101] checklist when describing exercise interventions. High-quality clinical trials directly comparing programs with varying numbers of resistance exercises while standardizing exercise dosage parameters across experimental conditions are needed to definitively answer our research question. Future work should also evaluate whether there is a relationship between number of resistance exercises and changes in muscle strength or objective functional performance measures in people with knee OA.

Our study has limitations. Given the limited information on exercise adherence that was available in included trials (Supplementary Table S1), we do not know whether the number of prescribed resistance exercises in our included RCTs accurately reflected the number of exercises that were actually performed by participants during the intervention. Furthermore, we do not know about other dosage parameters (eg, intensity), whether the program was progressed over time, or whether prescribed exercises were personalized to individual participants or if participants actually demonstrated muscle weakness. It is therefore unclear whether programs with a small number of resistance exercises were of a lower overall volume or load than programs with more resistance exercises. Our meta-regressions were modeled on a linear relationship, and we did not evaluate whether the relationship between number of exercise and outcomes might be nonlinear. However, our subgroup meta-analyses suggest that this was not the case. Finally, there were large variations in the participant characteristics of included trials, including their age, BMI, gender ratio, and severity of their symptoms at baseline.

In conclusion, there was no relationship between the number of different lower-limb resistance exercises prescribed in a program and change in knee pain or self-reported function. However, given that we were unable to account for all differences in program intensity, progression, and adherence, as well as the heterogeneity and overall low quality of included studies, our results should be interpreted with caution.

ACKNOWLEDGMENT

We thank Professor Karen Lamb (University of Melbourne) for her statistical advice.

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 Lawford 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. Cui A, Li H, Wang D, et al. Global, regional prevalence, incidence and risk factors of knee osteoarthritis in population-based studies. EClinicalMedicine 2020;29:100587.
2. Bannuru R, Osani M, Vaysbrot E, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27(11):1578–1589.
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;72:220–233.
4. Osteoarthritis in over 16s: diagnosis and management. National Institute for Health and Care Excellence; 2022.
5. Guideline for the management of knee and hip osteoarthritis. The Royal Australian College of General Practitioners. 2018. Accessed November 20, 2023. https://www.racgp.org.au/clinical-resources/clinical-guidelines/key-racgp-guidelines/view-all-racgp-guidelines/knee-and-hip-osteoarthritis
6. Moseng T, Vlieland TPMV, Battista S, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis: 2023 update. Ann Rheum Dis 2024;83(6):730–740.
7. Fransen M, McConnell S, Harmer AR, et al. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev 2015;1:CD004376.
8. 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.
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.
10. Muraki S, Akune T, Teraguchi M, et al. Quadriceps muscle strength, radiographic knee osteoarthritis and knee pain: the ROAD study. BMC Musculoskelet Disord 2015;16(1):305.
11. Vårbakken K, Lorås H, Nilsson KG, et al. Relative difference in muscle strength between patients with knee osteoarthritis and healthy controls when tested bilaterally and joint-inclusive: an exploratory cross-sectional study. BMC Musculoskelet Disord 2019;20(1):593.
12. Culvenor AG, Ruhdorfer A, Juhl C, et al. Knee extensor strength and risk of structural, symptomatic, and functional decline in knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2017;69(5):649–658.
13. Bandak E, Boesen M, Bliddal H, et al. Exercise-induced pain changes associate with changes in muscle perfusion in knee osteoarthritis: exploratory outcome analyses of a randomised controlled trial. BMC Musculoskelet Disord 2019;20(1):491.
14. Runhaar J, Luijsterburg P, Dekker J, et al. Identifying potential working mechanisms behind the positive effects of exercise therapy on pain and function in osteoarthritis; a systematic review. Osteoarthritis Cartilage 2015;23(7):1071–1082.
15. Hall M, Hinman RS, Wrigley TV, et al. Knee extensor strength gains mediate symptom improvement in knee osteoarthritis: secondary analysis of a randomised controlled trial. Osteoarthritis Cartilage 2018;26(4):495–500.
16. Runhaar J, Holden MA, Hattle M, et al; STEER OA Patient Advisory Group; OA Trial Bank Exercise Collaborative. Mechanisms of action of therapeutic exercise for knee and hip OA remain a black box phenomenon: an individual patient data mediation study with the OA Trial Bank. RMD Open 2023;9(3):e003220.
17. Liguori G, ed. ACSM's Guidelines for Exercise Testing and Prescription. Lippincott Williams & Wilkins; 2020.
18. Turner MN, Hernandez DO, Cade W, et al. The role of resistance training dosing on pain and physical function in individuals with knee osteoarthritis: a systematic review. Sports Health 2020;12(2):200–206.
19. Husted RS, Troelsen A, Husted H, et al. Knee-extensor strength, symptoms, and need for surgery after two, four, or six exercise sessions/week using a home-based one-exercise program: a randomized dose–response trial of knee-extensor resistance exercise in patients eligible for knee replacement (the QUADX-1 trial). Osteoarthritis Cartilage 2022;30(7):973–986.
20. de Zwart AH, Dekker J, Roorda LD, et al. High-intensity versus low-intensity resistance training in patients with knee osteoarthritis: a randomized controlled trial. Clin Rehabil 2022;36(7):952–967.
21. Torstensen TA, Østerås H, LoMartire R, et al. High-versus low-dose exercise therapy for knee osteoarthritis: a randomized controlled multicenter trial. Ann Intern Med 2023;176(2):154–165.
22. Hua J, Sun L, Teng Y. Effects of high-intensity strength training in adults with knee osteoarthritis: a systematic review and meta-analysis of randomized controlled trials. Am J Phys Med Rehabil 2023;102(4):292–299.
23. Marriott KA, Hall M, Maciukiewicz JM, et al. Are the effects of resistance exercise on pain and function in knee and hip osteoarthritis dependent on exercise volume, duration and adherence? Arthritis Care Res (Hoboken) 2024;76(6):821–830.
24. Fyfe JJ, Hamilton DL, Daly RM. Minimal-dose resistance training for improving muscle mass, strength, and function: a narrative review of current evidence and practical considerations. Sports Med 2022;52(3):463–479.
25. Fisher JP, Steele J. Heavier and lighter load resistance training to momentary failure produce similar increases in strength with differing degrees of discomfort. Muscle Nerve 2017;56(4):797–803.
26. Buckner SL, Dankel SJ, Mattocks KT, et al. The affective and behavioral responses to repeated "strength snacks". Physiol Int 2018;105(2):188–197.
27. Essery R, Geraghty AWA, Kirby S, et al. Predictors of adherence to home-based physical therapies: a systematic review. Disabil Rehabil 2017;39(6):519–534.
28. Westcott WL, Winett RA, Annesi JJ, et al. Prescribing physical activity: applying the ACSM protocols for exercise type, intensity, and duration across 3 training frequencies. Phys Sportsmed 2009;37(2):51–58.
29. Lawford BJ, Hall M, Hinman RS, et al. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev. 2024;12(12):CD004376.
30. Higgins JP, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions. Wiley; 2022.
31. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021;10(1):89.
32. Altman RD. Criteria for classification of clinical osteoarthritis. J Rheumatol Suppl 1991;27:10–12.

33. Zhang W, Doherty M, Peat G, et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010;69(3):483–489.
34. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 63. Wiley; 2022.
35. Deeks JJ, Higgins JP, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 63. Wiley; 2022.
36. Review Manager Web, version 3.1.2. The Cochrane Collaboration; 2020. https://revman.cochrane.org
37. Juhl C, Christensen R, Roos EM, et al. Impact of exercise type and dose on pain and disability in knee osteoarthritis: a systematic review and meta-regression analysis of randomized controlled trials. Arthritis Rheumatol 2014;66(3):622–636.
38. Egger M, Smith GD, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315(7109): 629–634.
39. Chen TW, Lin CW, Lee CL, et al. The efficacy of shock wave therapy in patients with knee osteoarthritis and popliteal cyamella. Kaohsiung J Med Sci 2014;30(7):362–370.
40. 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.
41. Foley A, Halbert J, Hewitt T, et al. Does hydrotherapy improve strength and physical function in patients with osteoarthritis—a randomised controlled trial comparing a gym based and a hydrotherapy based strengthening programme. Ann Rheum Dis 2003;62(12): 1162–1167.
42. Jan MH, Lin CH, Lin YF, et al. Effects of weight-bearing versus nonweight-bearing exercise on function, walking speed, and position sense in participants with knee osteoarthritis: a randomized controlled trial. Arch Phys Med Rehabil 2009;90(6):897–904.
43. Jorge RTB, de Souza MC, Chiari A, et al. Progressive resistance exercise in women with osteoarthritis of the knee: a randomized controlled trial. Clin Rehabil 2015;29(3):234–243.
44. Lin DH, Lin CHJ, Lin YF, et al. Efficacy of 2 non-weight-bearing interventions, proprioception training versus strength training, for patients with knee osteoarthritis: a randomized clinical trial. J Orthop Sports Phys Ther 2009;39(6):450–457.
45. Maurer BT, Stern AG, Kinossian B, et al. Osteoarthritis of the knee: isokinetic quadriceps exercise versus an educational intervention. Arch Phys Med Rehabil 1999;80(10):1293–1299.
46. 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.
47. Salli A, Sahin N, Baskent A, et al. The effect of two exercise programs on various functional outcome measures in patients with osteoarthritis of the knee: a randomized controlled clinical trial. Isokinet Exerc Sci 2010;18(4):201–209.
48. Samut G, Dinçer F, Özdemir O. The effect of isokinetic and aerobic exercises on serum interleukin-6 and tumor necrosis factor alpha levels, pain, and functional activity in patients with knee osteoarthritis. Mod Rheumatol 2015;25(6):919–924.
49. Schilke JM, Johnson GO, Housh TJ, et al. Effects of muscle-strength training on the functional status of patients with osteoarthritis of the knee joint. Nurs Res 1996;45(2):68–72.
50. Vincent KR, Vincent HK. Concentric and eccentric resistance training comparison on physical function and functional pain outcomes in knee osteoarthritis: a randomized controlled trial. Am J Phys Med Rehabil 2020;99(10):932–940.
51. Jan MH, Lin JJ, Liau JJ, et al. Investigation of clinical effects of high- and low-resistance training for patients with knee osteoarthritis: a randomized controlled trial. Phys Ther 2008;88(4):427–436.
52. Pazit L, Jeremy D, Nancy B, et al. Safety and feasibility of high speed resistance training with and without balance exercises for knee osteoarthritis: a pilot randomised controlled trial. Phys Ther Sport 2018; 34:154–163.
53. Baker KR, Nelson ME, Felson DT, et al. The efficacy of home based progressive strength training in older adults with knee osteoarthritis: a randomized controlled trial. J Rheumatol 2001;28(7):1655–1665.
54. de Almeida AC, Aily JB, Pedroso MG, et al. A periodized training attenuates thigh intermuscular fat and improves muscle quality in patients with knee osteoarthritis: results from a randomized controlled trial. Clin Rheumatol 2020;39(4):1265–1275.
55. Ettinger WH Jr, Burns R, Messier SP, et al. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). JAMA 1997 1997;277:25–31.
56. Imoto AM, Peccin MS, Trevisani VFM. Quadriceps strengthening exercises are effective in improving pain, function and quality of life in patients with osteoarthritis of the knee. Acta Ortop Bras 2012; 20(3):174–179.
57. Lai Z, Lee S, Chen Y, et al. Comparison of whole-body vibration training and quadriceps strength training on physical function and neuromuscular function of individuals with knee osteoarthritis: a randomised clinical trial. J Exerc Sci Fit 2021;19(3):150–157.
58. 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.
59. Bruce-Brand RA, Walls RJ, Ong JC, et al. Effects of home-based resistance training and neuromuscular electrical stimulation in knee osteoarthritis: a randomized controlled trial. BMC Musculoskelet Disord 2012;13:118.
60. Gür H, Çakın N, Akova B, et al. Concentric versus combined concentric-eccentric isokinetic training: effects on functional capacity and symptoms in patients with osteoarthrosis of the knee. Arch Phys Med Rehabil 2002;83(3):308–316.
61. Henriksen M, Klokker L, Graven-Nielsen T, et al. Association of exercise therapy and reduction of pain sensitivity in patients with knee osteoarthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2014;66(12):1836–1843.
62. Karadağ S, Taşci S, Doğan N, et al. Application of heat and a home exercise program for pain and function levels in patients with knee osteoarthritis: a randomized controlled trial. Int J Nurs Pract 2019; 25(5):e12772.
63. Krishnan K, Abadi FH, Zainudin FF, et al. Comparison of the effect of aquatic and Thera-band exercise on pain and quality of life in obese people with knee osteoarthritis. J Pain Manag 2021;14(3):221–229.
64. Lee EL, Jang MH, Lee BJ, et al. Home-based remote rehabilitation leads to superior outcomes for older women with knee osteoarthritis: a randomized controlled trial. J Am Med Dir Assoc 2023;24(10): 1555–1561.
65. Lim BW, Hinman RS, Wrigley TV, et al. Does knee malalignment mediate the effects of quadriceps strengthening on knee adduction moment, pain, and function in medial knee osteoarthritis? A randomized controlled trial. Arthritis Care Res (Hoboken) 2008;59(7): 943–951.
66. Mazloum V, Rabiei P, Rahnama N, et al. The comparison of the effectiveness of conventional therapeutic exercises and Pilates on pain and function in patients with knee osteoarthritis. Complement Ther Clin Pract 2018;31:343–348.

67. Topp R, Woolley S, Hornyak J III, et al. The effect of dynamic versus isometric resistance training on pain and functioning among adults with osteoarthritis of the knee. Arch Phys Med Rehabil 2002;83(9): 1187–1195.
68. Vassão PG, de Souza ACF, da Silveira Campos RM, et al. Effects of photobiomodulation and a physical exercise program on the expression of inflammatory and cartilage degradation biomarkers and functional capacity in women with knee osteoarthritis: a randomized blinded study. Adv Rheumatol 2021;61(1):62.
69. Wortley M, Zhang S, Paquette M, et al. Effects of resistance and Tai Ji training on mobility and symptoms in knee osteoarthritis patients. J Sport Health Sci 2013;2(4):209–214.
70. Anwer S, Alghadir A. Effect of isometric quadriceps exercise on muscle strength, pain, and function in patients with knee osteoarthritis: a randomized controlled study. J Phys Ther Sci 2014;26(5):745–748.
71. Assar S, Gandomi F, Mozafari M, et al. The effect of total resistance exercise vs. aquatic training on self-reported knee instability, pain, and stiffness in women with knee osteoarthritis: a randomized controlled trial. BMC Sports Sci Med Rehabil 2020;12(1):27.
72. Chang TF, Liou TH, Chen CH, et al. Effects of elastic-band exercise on lower-extremity function among female patients with osteoarthritis of the knee. Disabil Rehabil 2012;34(20):1727–1735.
73. Dinçer U, Arıba S, Saygın H, et al. The effects of closed kinetic chain exercise on articular cartilage morphology: myth or reality? A randomized controlled clinical trial. Turkiye Fiziksel Tip ve Rehabilitasyon Dergisi 2016;62(1):28–36.
74. Hsu YI, Chen YC, Lee CL, et al. Effects of diet control and telemedicine-based resistance exercise intervention on patients with obesity and knee osteoarthritis: a randomized control trial. Int J Environ Res Public Health 2021;18(15):7744.
75. Kholvadia A, Constantinou D, Gradidge PJL. Exploring the efficacy of low-level laser therapy and exercise for knee osteoarthritis. S Afr J Sports Med 2019;31(1):v31i1a6058.
76. Nahayatbin M, Ghasemi M, Rahimi A, et al. The effects of routine physiotherapy alone and in combination with either tai chi or closed kinetic chain exercises on knee osteoarthritis: a comparative clinical trial study. Iranian Red Crescent Med J 2018;20(4):e62600.
77. O'Reilly SC, Muir KR, Doherty M. Effectiveness of home exercise on pain and disability from osteoarthritis of the knee: a randomised controlled trial. Ann Rheum Dis 1999;58:15–19.
78. Oh SL, Kim DY, Bae JH, et al. Effects of rural community-based integrated exercise and health education programs on the mobility function of older adults with knee osteoarthritis. Aging Clin Exp Res 2021; 33(11):3005–3014.
79. Rafiq MT, Hamid MSA, Hafiz E. Effect of progressive resistance strength training on body mass index, quality of life and functional capacity in knee osteoarthritis: a randomized controlled trial. J Multidiscip Healthc 2021:2161–2168.
80. Rafiq MT, Hamid MSA, Hafiz E. The effect of rehabilitation protocol using mobile health in overweight and obese patients with knee osteoarthritis: a clinical trial. Adv Rheumatol 2021;61(1):63.
81. Rafiq MT, Hamid MSA, Hafiz E. Short-term effects of strengthening exercises of the lower limb rehabilitation protocol on pain, stiffness, physical function, and body mass index among knee osteoarthritis participants who were overweight or obese: a clinical trial. ScientificWorldJournal 2021;2021:6672274.
82. Doss SSD, Rekha K, Prathap S. Effects of non weight bearing strength training for knee osteoarthritis. Int J Res Pharm Sci 2014; 5(3):188–193.
83. Rêgo TAM, de Lima Ferreira AP, Villela DW, et al. Effects of mat Pilates on older adult women with knee osteoarthritis: a randomized controlled trial. J Bodywork Mov Ther 2023;33:136–141.
84. Lawford BJ, Hinman RS, Spiers L, et al. Does higher compliance with American College of Sports Medicine exercise prescription guidelines influence exercise outcomes in knee osteoarthritis? A systematic review with meta-analysis. Arthritis Care Res (Hoboken) 2025;77(4): 460–474.
85. Nuzzo JL, Pinto MD, Kirk BJ, et al. Resistance exercise minimal dose strategies for increasing muscle strength in the general population: an overview. Sports Med 2024;54(5):1139–1162.
86. 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 Cart Open 2024;6(1):100431.
87. Lawford B, Hinman RS, McManus FJ, et al. How does exercise, with and without diet, improve pain and function in knee osteoarthritis? A secondary analysis of a randomized controlled trial exploring potential mediators of effects. Arthritis Care Res (Hoboken) 2023;Nov; 75(11):2316–2327.
88. Bandak E, Christensen R, Overgaard A, et al. Exercise and education versus saline injections for knee osteoarthritis: a randomised controlled equivalence trial. Ann Rheum Dis 2022;81(4):537–543.
89. Henriksen M, Christensen R, Kristensen LE, et al. Exercise and education vs intra-articular saline for knee osteoarthritis: a 1-year follow-up of a randomized trial. Osteoarthritis Cartilage 2023;31(5):627–635.
90. Englund M, Turkiewicz A. The emperor's new clothes? Osteoarthritis Cartilage 2023;31(5):549–551.
91. Medina-Mirapeix F, Escolar-Reina P, Gascón-Cánovas JJ, et al. Predictive factors of adherence to frequency and duration components in home exercise programs for neck and low back pain: an observational study. BMC Musculoskelet Disord 2009;10:155.
92. Palazzo C, Klinger E, Dorner V, et al. Barriers to home-based exercise program adherence with chronic low back pain: patient expectations regarding new technologies. Ann Phys Rehabil Med 2016; 59(2):107–113.
93. Dickson C, de Zoete RMJ, Berryman C, et al. Patient-related barriers and enablers to the implementation of high-value physiotherapy for chronic pain: a systematic review. Pain Med 2024;25(2):104–115.
94. Dobson F, Bennell KL, French SD, et al. Barriers and facilitators to exercise participation in people with hip and/or knee osteoarthritis: synthesis of the literature using behavior change theory. Am J Phys Med Rehabil 2016;95(5):372–389.
95. Henry KD, Rosemond C, Eckert LB. Effect of number of home exercises on compliance and performance in adults over 65 years of age. Phys Ther 1999;79(3):270–277.
96. Husted RS, Bandholm T, Rathleff MS, et al. Perceived facilitators and barriers among physical therapists and orthopedic surgeons to preoperative home-based exercise with one exercise-only in patients eligible for knee replacement: a qualitative interview study nested in the QUADX-1 trial. PLoS One 2020;15(10):e0241175.
97. Hurst C, Robinson SM, Witham MD, et al. Resistance exercise as a treatment for sarcopenia: prescription and delivery. Age Ageing 2022;51(2):afac003.
98. Hurst C, Sayer AA. Improving muscle strength and physical function in older people living with sarcopenia and physical frailty: not all exercise is created equal. J R Coll Physicians Edinb 2022;52(2):166–171.
99. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869.
100. Slade SC, Dionne CE, Underwood M, et al. Consensus on Exercise Reporting Template (CERT): explanation and elaboration statement. Br J Sports Med 2016;50(23):1428–1437.
101. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014;348:g1687.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 604–613
DOI 10.1002/acr.25466

American College of Rheumatology
Empowering Rheumatology Professionals

Incidence of and Risk of Mortality After Hip Fractures in Rheumatoid Arthritis Relative to the General Population

C. Allyson Jones,[1] Pierre Guy,[2] Hui Xie,[3] Eric C. Sayre,[4] Kai Zhao,[3] and Diane Lacaille[5]

Objective. Osteoporosis, a known complication of rheumatoid arthritis (RA), increases the risk of hip fracture, which is associated with high morbidity and mortality. Fracture risk estimates in patients with RA treated with contemporary treatment strategies are lacking. The objectives were (1) estimate age-specific and sex-specific incidence rates and compare the risk of hip fractures in RA relative to age-matched and sex-matched general population controls, and (2) compare the risk of all-cause mortality in RA and general population controls after hip fracture.

Methods. A longitudinal study of a population-based incident cohort of patients with RA diagnosed between 1997 and 2009, followed until 2014, with age-matched and sex-matched controls from the general population of British Columbia, using administrative health data. Hip fracture outcomes (International Classification of Diseases, Ninth Edition, Clinical Modification [ICD-9-CM] codes 820.0 or 820.2; ICD-10-Canada code S72.0 to S72.2) and mortality at predefined intervals after fracture (in hospital, 90 days, 1-year, 5-year) were identified. Hip fracture incidence rates for RA and controls, and incidence rate ratios (IRRs), were calculated. Cox proportional hazards models compared hip fracture and mortality risk in RA versus controls; logistic regression compared in-hospital mortality risk.

Results. Overall, 1,314 hip fractures over 360,521 person-years were identified in 37,616 individuals with RA and 2,083 over 732,249 person-years in 75,213 controls, yielding a 28% greater fracture risk in RA (IRR 1.28 [95% confidence interval 1.20–1.37]). Mean age at time of fracture was slightly younger for RA than controls (79.6 ± 10.8 vs 81.6 ± 9.3 years). Postfracture mortality risk at one-year and five-years did not differ between RA and general population controls. Results were similar in a sensitivity analysis including only individuals with RA who received disease-modifying antirheumatic drugs.

Conclusion. People with RA had a greater risk of hip fractures, but no greater risk of mortality post fracture, than the general population. The relative risk of hip fractures observed was not as high as previously reported, likely reflecting better treatment of inflammation and management of osteoporosis and its risk factors.

INTRODUCTION

Rheumatoid arthritis (RA) is associated with fragility fractures and is the only disease that is specifically identified as a risk factor for fracture within the Fracture Risk Assessment Tool (FRAX).[1] Although vertebral fractures account for half of fragility fractures in RA,[2] hip fracture is a public health burden because of the high morbidity, mortality, disability, and socioeconomic costs.[3,4] Hip fractures have been extensively studied in older patient populations, yet a paucity of literature has examined hip fracture in RA despite osteoporosis affecting slightly more than 25% of the RA population.[5] The burden of hip fracture cannot be underestimated. Within the general population, temporal and geographic variations of incidence rates exist worldwide with an estimated incidence rate of 14.2 (95% confidence interval [95% CI] 11.1–18.1) per 1,000,000 in 2019.[6] In Canada, the national age and sex standardized incidence rate is 15.78 (95% CI 15.72–15.83) per 1,000,000.[7] The one-year mortality rates after

Supported by the Canadian Institute of Health Research (grant MOP-130480) and the Canadian Initiative for Outcomes in Rheumatology Care. Dr Lacaille's work was supported by the Mary Pack Chair in Rheumatology Research from the University of British Columbia and The Arthritis Society of Canada.

[1]C. Allyson Jones, PT, PhD: University of Alberta, Edmonton, Alberta, and Arthritis Research Canada, Vancouver, British Columbia, Canada; [2]Pierre Guy, MD, MBA, FRCSC: University of British Columbia, Vancouver, British Columbia, Canada; [3]Hui Xie, PhD, Kai Zhao, MSc: Arthritis Research Canada, Vancouver, and Simon Fraser University, Burnaby, British Columbia, Canada; [4]Eric C. Sayre, PhD: Arthritis Research Canada, Vancouver, British Columbia, Canada; [5]Diane Lacaille, MD, FRCPC, MHSc: Arthritis Research Canada and University of British Columbia, Vancouver, British Columbia, Canada.

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.25466).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25466.

Address correspondence via email to Diane Lacaille, MD, FRCPC, MHSc, at dlacaille@arthritisresearch.ca.

Submitted for publication January 21, 2024; accepted in revised form November 5, 2024.

SIGNIFICANCE & INNOVATIONS

- Rheumatoid arthritis (RA) is a risk factor for fragility fractures including hip fractures.
- Hip fracture is a public health burden with high morbidity and mortality risk.
- Relative risk of hip fractures observed was not as high as historically reported, likely reflecting better treatment of inflammation and management of osteoporosis and its risk factors.
- Nonetheless, the risk of hip fracture remains higher than in general population controls of the same age and sex, despite advances in treatment.
- Mortality after hip fracture did not differ significantly from that of general population controls for the same age and sex.
- Given the persistently increased risk of hip fractures despite advances in treatment of inflammation, and the high morbidity and mortality following hip fractures, fall prevention programs and other primary prevention strategies targeting RA are needed.

a hip fracture are high, ranging up to 35%.[7,8] The excess mortality rate is at least double for age-matched population norms with the greatest mortality risk within the first three to six months after the index hip fracture.[9,10] Recovery after hip fracture in the general population is long-term, lasting two or more years,[11] with many patients unable to return to independent community living.[4,11,12]

A recent systematic review reported the pooled incidence rate of hip fractures in RA of 4.33 (95% CI 2.26–8.27) per 1,000 person-years.[2] Moreover, the secular trend of hip fracture in RA populations is increasing,[13] unlike the static or decreasing trends reported in the general population.[3,7] This epidemiologic pattern of hip fracture was illustrated in a Spanish cohort, in that patients with RA tended to be younger than the general population, and the incidence of hip fracture in the RA cohort increased over the 17 year observation period.[13] Mortality is high after a hip fracture, yet mortality is not commonly reported in RA cohorts with hip fractures.[13–15] There is consistent evidence, however, that mortality post hip fracture is greater in men than women regardless of whether a patient has RA or not.[7,15]

Although hip fracture in RA has been examined across different geographic regions and ethnicities, several studies have evaluated risk in prevalent RA cohorts, which can provide biased risk estimates.[2,16] The methodology is inconsistent in terms of study designs, case ascertainment, and reporting fractures, all of which may account for the varied incidence rates reported in the literature. Assembling a population-based cohort of incident RA followed from diagnosis will provide a reliable estimate of hip fracture risk relative to the general population. The objectives of this study are two-fold: (1) to estimate age-specific and sex-specific incidence rates and compare the risk of hip fractures in RA relative to age-matched and sex-matched general population controls, and (2) to compare the risk of all-cause mortality in RA and general population controls after a hip fracture.

METHODS

Study design and sample. A longitudinal study of a population-based incident RA cohort with age-matched and sex-matched general population controls using administrative health data for the province of British Columbia (BC), Canada.

Incident RA cohort. An incident cohort of persons with RA was identified using a previously published RA definition.[17] Using physician billing data, all incident patients with RA in BC who first met RA criteria between January 1, 1997, and December 31, 2009, were identified and followed until December 31, 2014. The case definition included having a minimum of two physician visits more than two months apart, within a five-year period, with International Classification of Diseases, Ninth Edition (ICD-9) codes for RA (714.X) and/or the Tenth version (ICD-10-Canada: M05.X–M06.X). The rationale for two visits rather than a single visit was to avoid cases in which an initial impression of RA, or a visit to evaluate for possible RA, was not confirmed on later visits. The reasoning for the two physician visits being more than two months apart was to exclude transient inflammatory arthritis. Individuals were excluded if, over a five-year period after their second RA visit (ie, RA index date), they had at least two subsequent visits, on two different days, with the same diagnostic code for another form of inflammatory arthritis (systemic lupus erythematosus, other connective tissue diseases, psoriatic arthritis, ankylosing spondylitis, and other spondyloarthropathies); or if a diagnosis of RA by a nonrheumatologist was never confirmed during subsequent visits to a rheumatologist. These criteria have been validated in a subsample who participated in a RA survey, using opinion of an independent rheumatologist reviewing medical records from their treating physicians as gold standard, yielding a positive predictive value (PPV) of 0.82.[18] Date of the second ICD code for RA diagnosis was used as the RA index date. To identify incident cases, a run-in period of seven years was used (selected to have the longest run-in period possible, because the earliest data available were from 1990 onward). This ensured prevalent cases of RA who moved to BC were not erroneously identified as incident RA cases, by excluding individuals with RA with less than seven years of available data in Medical Service Plan (MSP) registry prior to their first RA visit.

Matched general population controls. A random control sample without any diagnosis of RA or other inflammatory arthritis was assembled from the general population (BC population was 3.9–4.4 million in 1997–2009[19]) using the same administrative databases as for the RA cohort. Controls were matched to individuals with RA using a 2:1 ratio on birth year, sex, and calendar year of MSP enrollment, and were assigned the same RA index date as the corresponding RA case.

Exclusion criteria applied to the RA cohort and general population controls included hip fractures, pathologic fractures, or Paget's disease occurring at any time prior to RA index date (or index date in controls), using all available data from 1990 onward. To minimize the possibility of missing prior hip fractures and inadvertently classifying a second hip fracture as an incident hip fracture in individuals who moved to BC, a run-in period of seven years was used, whereby individuals with less than seven years of data prior to (RA) index date were excluded.

Data sources. Complete health information for physician visits, medications dispensed, and hospitalizations were obtained from administrative databases of the Ministry of Health, through Population Data BC until December 2014. Within this public health care system, all persons are guaranteed universal coverage for physician visits and hospital and medical services, including surgical treatment of hip fracture. Our specific data sources are listed below:

1. Canadian Institute of Health Information Hospital Separation Abstracts (January 1990 to December 2014)[20] include up to 25 diagnostic codes per hospitalization using full five-digit ICD-9 and/or ICD-10 codes, representing either the reason for admission or complications during hospitalization, hospital admission and discharge dates, and hospital transfers. One hundred percent of hospitalizations in BC are captured through this system.
2. MSP File (January 1990 to December 2014)[21] contains physician claims data used for reimbursement of physician visits, under fee-for-service billings. Each claim contains a single diagnosis representing the reason for the visit based on ICD-9 diagnostic codes. Approximately 95% of physician-patient episodes of care in BC are reimbursed through the BC Ministry of Health fee-for-service.
3. MSP Consolidation File (January 1990 to December 2014)[21] includes demographic information for each person registered with the provincial plan, such as age, sex, postal code (first three digits) used to determine rural versus urban residence, neighborhood income quintile and local health area, and registration data.
4. BC Vital Statistics (January 1996 to December 2014)[22] were obtained on the date of death and primary cause of death (ICD-9 or ICD-10) from information provided on death certificates.
5. Pharmanet Database (January 1996 to December 2014)[23] provides information on all prescriptions dispensed by pharmacies in BC, regardless of source of funding.

Hip fracture outcomes were identified using ICD9-CM codes 820.0 or 820.2 and ICD10-Canada code S72.0 to S72.2 placed in any position in hospitalization data, representing either the reason for admission or a complication occurring during hospitalization.[24,25] The accuracy of this algorithm to identify hip fractures from administrative data has been validated when used with hospital data (sensitivity 83%–97%; PPV 86%–98%).[26] This approach also ensured that we captured hip fractures that may have occurred while in hospital for other reasons.[27] Pathologic fractures (metastases; disorders of bone such as Paget's disease, fractures of other or unspecified femur location or acetabular-pelvic fractures) were not included.

Mortality outcomes. All-cause mortality was assessed at predetermined time points: in hospital, 90-day, 1-year and 5-year post fracture to capture short and long-term complications of hip fractures. Although not all deaths post hip fracture may be directly attributable to fractures,[28] we opted to evaluate all-cause mortality to capture the broadest impact of hip fracture on people's health that can contribute to increased mortality. The definition of an episode of care as outlined by Sheehan et al was used to assess in-hospital mortality, to take into consideration hospital transfers.[29]

Covariates. Baseline covariates were assessed within 12 months prior to RA index date in the analyses evaluating risk of hip fracture, and within 12 months prior to hip fracture date in the analyses evaluating mortality risk post hip fracture. Covariates included age, sex, social economic status (SES) based on the first three postal code digits and neighborhood income quintile, urban/rural area, the Romano modification of the Charlson comorbidity index (excluding RA from comorbidities) for use with administrative data (cardiovascular disease including coronary artery disease, congestive heart failure, peripheral vascular disease, venous thromboembolism; respiratory conditions; dementia; depression; malignancy; cerebrovascular accident; chronic liver disease/cirrhosis; alcoholism; chronic kidney disease),[30–32] and hospital size for postfracture mortality analyses. A single diagnostic code (ICD-9/10) in hospitalization data and physician service claims were used to identify comorbidities (see Supplementary Table 1).

Statistical analysis. Descriptive statistics (ie, means, SDs, or frequency counts and percentages) were computed for relevant variables in the RA cohort and general population controls. Incidence rates of hip fractures were calculated for the RA cohort and general population controls, and risk of hip fracture in RA relative to the general population was estimated using crude incidence rate ratios (IRR) with 95% CIs, and adjusted hazard ratios (aHRs) from Cox's proportional hazard models (PHMs) adjusted for age, sex, and Romano Charlson (excluding RA). For each individual, follow-up started at (RA) index date and ended at occurrence of first hip fracture, with censoring of follow-up at time of death, out-of-province migration, or end of follow-up (December 2014), whichever occurred first. Covariates were selected for inclusion as potential confounders in the final PHMs according to a purposeful selection algorithm using a 5% threshold for inclusion. The algorithm proceeded by entering potential confounders one at a time into the PHM, and assessing the relative change in the

HR estimating fracture risk in RA relative to the general population, compared with the previous model.[33] The algorithm stopped when no additional variables had a ≥5% impact on the HR for RA.[33,34] Alternative models were fit based on the subdistribution hazards to accommodate for the competing risk of death.[35] By comparing results from the subdistribution hazards regression model and Cox's PHMs, the two primary statistical models for competing risk analysis, we evaluated the robustness of adjusted HR estimates across different modeling methods. We also used the Fine-Gray method[36] to compute the cumulative incidence function (CIF) of first hip fracture, while accounting for competing risks of death because of causes unrelated to hip fracture. Gray's Test[37] was used for comparing the CIFs between the two groups.

Mortality post hip fracture was analyzed via Cox PHMs in all RA and general population patients who sustained a hip fracture. For each individual, follow-up started at fracture date and ended at occurrence of death, with censoring of follow-up at out-of-province migration or end of follow-up (December 2014), whichever occurred first. We fit unadjusted models, models adjusted for age and sex, and models adjusted for age, sex, SES, rural/urban, hospital size, and fracture type. With the sensitivity analyses, we also fit models adjusted for age, sex, SES, rural/urban area, hospital size, fracture type, and the following comorbidities influencing risk of death: Romano Charlson (excluding RA), cardiovascular disease, chronic obstructive pulmonary disease (COPD), malignancy, dementia, alcoholism, diabetes, hyperlipidemia, Hormonal Replacement Therapy, and anticoagulants. Although matching variables age and sex were not confounders, they are known strong predictors of the outcomes, and adjusting for them increases the power to detect between group differences. Each model estimated the effects of RA (vs general population) on mortality according to time since fracture: <90 days, 90 days to 1 year, and 1 to 5 years post fracture. In addition, we computed cumulative all-cause mortality, using product limit estimates, at 90-day, 1-year, and 5-year post fracture, stratified by RA status. In-hospital mortality (binary yes/no variable) was estimated using logistic regression models. Sensitivity analyses were also conducted limiting the sample to individuals with RA who had received disease-modifying antirheumatic drugs (DMARDs) and their age-matched and sex-matched controls to ensure that hip fracture and mortality risks relative to the general population controls were not affected by our RA definition, which did not require specific RA treatment. Analyses were performed using SAS version 9.4 (SAS Institute Inc). Ethics approval was obtained from the Health Research Ethics Review Board at the University of British Columbia (H00-80305) and University of Alberta (PRO 00071977).

RESULTS

A total of 37,616 persons with RA and 75,213 age/sex-matched general population controls were followed from January 1, 1997, to December 31, 2014. The sample included 66% females, and the mean (SD) age at RA diagnosis was 57.3 (16.6) years (Table 1). Within the RA cohort, hydroxychloroquine (30.0%) and methotrexate (26.3%) were the most commonly dispensed DMARDs, followed by sulfasalazine (13%) and leflunomide (6%), whereas 6.2% received biologics. Over 360,521 person-years of follow-up, 1,314 hip fractures occurred in the RA cohort (mean [SD] follow-up time: 9.6 [4.3] years), whereas 2,083 hip fractures were reported over 732,249 person-years of follow-up for the matched general population controls (mean [SD] follow-up time: 9.7 [4.3] years), yielding incident rates per 1,000 person-years of follow-up of 3.6 (95% CI 3.4–3.8) in RA and 2.8 (95% CI 2.7–3.0) in general population controls. The mean (SD) age at time of hip fracture was slightly younger for the RA cohort than the general population controls (79.5 [10.8] vs 81.6 [9.3] years; $P < 0.001$). At the time of hip fracture, the RA cohort had more comorbidities than controls, as reflected by a higher mean (SD) Romano comorbidity score (RA: 1.3 [1.8] vs control: 1.08 [1.1]; $P = 0.002$). The most frequent comorbidity reported in both groups at the time of hip fracture was cardiovascular diseases (RA 62%; control 63%), whereas chronic obstructive lung conditions (COPD, emphysema, chronic bronchitis, asthma) (18%) was the next most reported comorbidity for the RA group and dementia (23%) for the control group.

Fracture characteristics and surgical management. No differences were seen between the two groups in the type of hip fracture; slightly less than one-half of both groups had transcervical hip fractures (Table 1). Approximately 23% in both groups were transferred to another hospital for surgery, and 7% in both groups did not receive surgery after admission for hip fracture. The median (25th, 75th percentile) time from hospital admission to surgery was 1.0 (1.0, 2.0) days for both groups. More people with RA resided in rural locales (16.8%) than general population controls (11.6%) at the time of hip fracture, with a greater number of patients with RA hospitalized in smaller hospitals (Table 1).

Risk of hip fracture. The crude IRR for hip fractures was 1.28 (95% CI 1.20–1.37) (Figure 1), indicating that individuals with RA had a 28% higher risk of hip fractures than general population controls. The IRR was slightly higher in males than females (1.45 [95% CI 1.25–1.69] vs 1.24 [95% CI 1.14–1.34]), although the difference was not statistically significant. The IRR for hip fracture decreased with increasing age. Individuals diagnosed with RA before the age of 60 had a greater than two-fold increase in risk of hip fracture, whereas the increased risk was less than two-fold for individuals diagnosed after the age of 60. The risk did not differ significantly from general population controls in individuals diagnosed after the age of 80 (Figure 1).

Multivariable Cox PHMs estimating the risk of incident hip fracture in RA relative to general population controls, age-adjusted and sex-adjusted HR was 1.28 (95% CI 1.19–1.37) (Table 2). In the final fully adjusted model, the Romano comorbidity score was the only additional covariate selected. The Cox PHM

Table 1. Characteristics of incident RA cohort and general population controls*

	RA	Controls	*P* value
n	37,616	75,213	—
Person-years of follow-up	360,521	732,249	—
Female, n (%)	24,987 (66)	49,880 (66)	0.717
Age at index date,[a] mean (SD), y	57.3 (16.6)	57.2 (16.6)	0.771
Hip fractures, n (%)	1,314 (3.5)	2,083 (2.8)	**<0.001**
Incidence rate per 1,000 person-years (95% CI)	3.6 (3.4–3.8)	2.8 (2.7–3.0)	**<0.001**
Hip fracture characteristics			
Age at hip fracture, mean (SD), y	79.5 (10.8)	81.6 (9.3)	**<0.001**
Female, n (%)	1,019 (77.6)	1,667 (80.0)	0.084
Type of hip fracture, n (%)			0.899
Cervical	626 (47.6)	997 (47.9)	—
Trochanteric[b]	688 (52.4)	1,086 (52.1)	—
Type of surgical fixation, n (%)			0.976
Total hip arthroplasty	46 (3.5)	70 (3.4)	—
Hemiarthroplasty	143 (10.9)	222 (10.7)	—
Internal fixation	1,031 (78.5)	1,635 (78.5)	—
No surgery	94 (7.2)	156 (7.5)	—
Chronic conditions[c]			
Cardiovascular disease	816 (62.1)	1,308 (62.8)	0.684
Respiratory conditions	243 (18.5)	306 (14.7)	**0.003**
Dementia	230 (17.5)	475 (22.8)	**<0.001**
Depression	200 (15.2)	292 (14.0)	0.332
Cancer and malignancy	143 (10.9)	202 (9.7)	0.265
Cerebrovascular accident	92 (7.0)	149 (7.2)	0.867
Chronic liver disease/cirrhosis	37 (2.8)	38 (1.8)	0.055
Romano score,[d] mean (SD)	1.3 (1.8)	1.1 (1.5)	**0.002**
LOS, median (25Q–75Q), d	15 (8–31)	16 (8–33)	0.635
Hospital size for hip surgery			**<0.001**
<50 beds	100 (7.6)	127 (6.1)	**—**
50–199 beds	474 (36.1)	613 (29.4)	**—**
200+ beds	740 (56.3)	1,343 (64.5)	**—**
Hospital transfer, n (%)	306 (23.3)	484 (23.2)	0.972
Residence locale at hip fracture time			**<0.001**
Rural	221 (16.8)	241 (11.6)	**—**
Urban	1,093 (83.2)	1,842 (88.4)	**—**
Socioeconomic status quintile of neighborhood at hip fracture time			0.767
1, lowest	351 (26.7)	547 (26.3)	—
2	273 (20.8)	409 (19.6)	—
3	261 (19.9)	408 (19.6)	—
4	221 (16.8)	384 (18.4)	—
5, highest	208 (15.8)	335 (16.1)	—

* Unless otherwise indicated, values represent n (%). The bolded *P* values denote statistical significance. 95% CI, 95% confidence interval; LOS, length of stay; Q, quarter; RA, rheumatoid arthritis.
[a] Age at index date of RA diagnosis; control group was assigned the same index date as the corresponding RA cases.
[b] Pertrochanteric or subtrochanteric hip fracture.
[c] Chronic conditions and Romano scores evaluated within 12 months prior to the date of hip fracture.
[d] Romano adaptation of Charlson comorbidity index developed for administrative health data, excluding RA as a comorbidity.

adjusted for age, sex, and Romano score estimated that persons with RA had a 27% greater risk of hip fracture than general population controls (aHR 1.27 [95% CI 1.18–1.36]); the relative risk was greater for males than females (aHR 1.44 [95% CI 1.24–1.67] vs aHR 1.23 [95% CI 1.13–1.33]) (Table 2), although the difference was not statistically significant. Cox PHMs stratified by age showed similar trends in risk relative to the general population according to age as described above for the IRR results. Alternative analysis accounting for competing events using different approaches yielded similar results attesting to the robustness of our findings. Specifically, when looking at the cumulative incidence of hip fracture, accounting for the competing risk of death, persons with RA had a statistically significant higher risk and shorter time to acquire a first hip fracture than general population controls (Gray's test *P* value <0.001) (Figure 2). Subdistribution PHMs accounting for the competing risks of death did not produce substantially different estimates compared with the standard cause-specific PHMs (Table 2).

Risk of mortality post hip fracture. A total of 715 deaths post hip fracture were observed in the RA cohort and 1,224 in the general population controls (Table 3). Of the

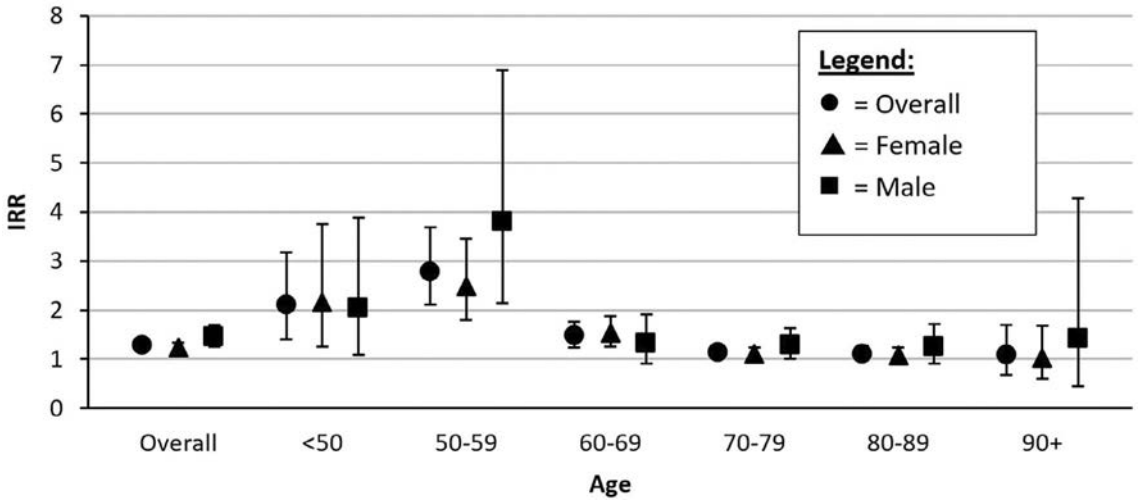


Figure 1. IRRs of hip fracture in rheumatoid arthritis versus controls, by age and sex. IRR, incidence rate ratio.

Table 2. PHM and sHR estimating risk of hip fracture in RA relative to general population controls*

	Overall		Females		Males	
Cox PHMs	cHR (95% CI)	sHR (95% CI)	cHR (95% CI)	sHR (95% CI)	cHR (95% CI)	sHR (95% CI)
Unadjusted models	1.29 (1.20– 1.38)	1.27 (1.18– 1.36)	1.24 (1.15–1.34)	1.22 (1.13–1.32)	1.46 (1.26–1.69)	1.43 (1.23–1.66)
Age- and sex-adjusted models	1.28 (1.19–1.37)	1.26 (1.17–1.35)	1.23 (1.14–1.33)	1.22 (1.12–1.32)	1.46 (1.26–1.69)	1.43 (1.23–1.66)
Fully adjusted models[a]	1.27 (1.18–1.36)	1.26 (1.17–1.35)	1.23 (1.13–1.33)	1.23 (1.12–1.32)	1.44 (1.24–1.67)	1.43 (1.23–1.66)
Fully adjusted models stratified by age						
<50 y	2.06 (1.39–3.04)	2.03 (1.37–3.00)	2.08 (1.24–3.50)	2.09 (1.25–3.49)	2.00 (1.11–3.62)	1.95 (1.07–3.57)
50–59 y	2.72 (2.07–3.56)	2.69 (2.05–3.52)	2.47 (1.80–3.37)	2.44 (1.78–3.33)	3.57 (2.06–6.16)	3.54 (2.05–6.12)
60–69 y	1.47 (1.23–1.74)	1.44 (1.21–1.71)	1.51 (1.24–1.84)	1.49 (1.23–1.82)	1.32 (0.92–1.90)	1.28 (0.90–1.83)
70–79 y	1.14 (1.03–1.27)	1.13 (1.01–1.26)	1.11 (0.98–1.25)	1.09 (0.97–1.23)	1.29 (1.02–1.63)	1.28 (1.01–1.61)
80–89 y	1.10 (0.96–1.25)	1.11 (0.98–1.27)	1.07 (0.93–1.23)	1.09 (0.94–1.25)	1.24 (0.91–1.69)	1.25 (0.92–1.71)
90+ y	1.12 (0.72–1.73)	1.03 (0.67–1.59)	1.06 (0.65–1.74)	0.98 (0.61–1.59)	1.39 (0.51–3.75)	1.27 (0.46–3.47)

* 95% CI, 95% confidence interval; cHR, cause-specific hazard ratio; PHM, proportional hazard model; RA, rheumatoid arthritis; sHR, subdistribution hazard ratio, which accounts for competing risk of death.
[a] Adjusted for age, sex, and Romano comorbidity score (excluding RA).

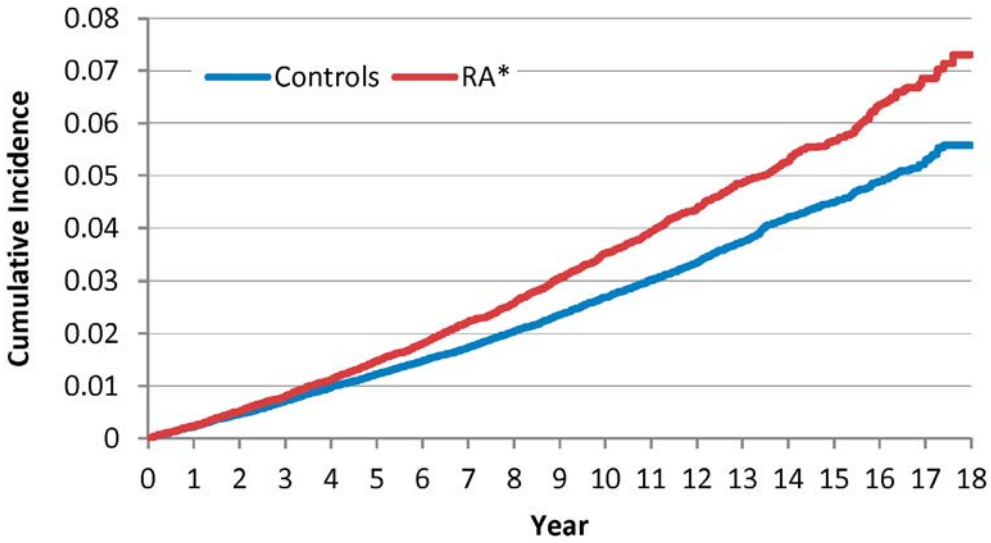


Figure 2. Cumulative incidence functions of hip fracture among RA and controls, adjusting for death as competing events. *Significant difference *P* value <0.001 (Gray's test). RA, rheumatoid arthritis.

Table 3. All-cause mortality after hip fracture*

Characteristics of people who died	RA	Controls	*P* value
Number of deaths	715	1,224	—
Age at death, mean (SD), y	84.8 (8.9)	86.2 (7.6)	**0.009**
Females, n (%)	537 (75.1)	960 (78.4)	0.092
Rural, n (%)	123 (17.2)	133 (10.9)	**<0.001**
Hospital LOS, median (IQR)	18 (9–36)	19 (9–37)	0.782
Socioeconomic status quintile of neighborhood at time of hip fracture			0.125
1 lowest	197 (27.6)	338 (27.6)	—
2	161 (22.5)	230 (18.8)	—
3	116 (16.2)	248 (20.3)	—
4	131 (18.3)	215 (17.6)	—
5 highest	110 (15.4)	193 (15.8)	—
Time from fracture to death, median (IQR), d	764 (156–1,518)	631 (107–1,415)	**0.035**
Cumulative number (%) deaths			—
In-hospital death	94 (13.1)	182 (14.9)	0.295
Death ≤90 d	137 (19.2)	284 (23.2)	**0.006**[a]
Death ≤1 y	252 (35.2)	470 (38.4)	0.584[a]
Death ≤5 y	585 (81.8)	1,027 (83.9)	**0.021**[a]

* The bolded *P* values denote statistical significance. IQR, interquartile range; LOS, length of stay; RA, rheumatoid arthritis.
[a] *P* values are from corresponding event time models.

patients who did not receive surgery, in-hospital mortality for individuals with RA was 26.6% (n = 25) and 29.5% (n = 46) for general population controls. Mean age at death was slightly younger for the RA cases than general population controls (84.8 [8.9] years vs 86.2 [7.6] years; $P = 0.009$). The 90-day mortality rate was lower in individuals with RA than general population controls (456.6 deaths per 1,000 person-years vs 610.4 deaths per 1,000 person-years; $P = 0.006$) (Table 3). The Cox PHM analyses adjusted for age, sex, and other sociodemographic factors potentially influencing postfracture mortality (ie, SES, rural/urban location, hospital size, and fracture type); patients with RA had a 20% lower 90-day mortality risk relative to general population controls (aHR 0.80 [95% CI 0.65–0.98]) (Table 4). Sensitivity analyses in which a third model adjusted for the covariates included in adjusted model 2, as well as for comorbidities potentially influencing mortality risk, yielded similar results (adjusted model 3, Table 4). Otherwise, cumulative mortality rates and mortality risks over different time periods post hip fractures did not differ significantly between patients with RA and general population controls who sustained hip fractures (Tables 4).

Sensitivity analyses conducted on the subgroup of patients with RA who received DMARDs did not reveal substantial differences in results from the main analyses (see Supplementary Tables 2–5). The fully adjusted Cox PHM estimated that persons with RA who received DMARDs had a 41% greater risk of hip fracture than general population controls (aHR 1.41 [95% CI 1.26–1.59]), a risk only slightly higher than in the entire RA cohort (aHR 1.27 [95% CI 1.18–1.36]) (see Supplementary Table 3). Estimates of mortality post hip fracture were very similar to those of the entire RA cohort, and mortality risk post hip fracture did not differ significantly between patients with RA receiving DMARDs and general population controls (see Supplementary Table 5).

Table 4. Mortality risk in RA relative to general population controls after hip fracture*

Cumulative mortality	Mortality rate RA (per 1,000 PY)	Mortality rate controls (per 1,000 PY)	Period-specific HR	Unadjusted HR (95% CI)	Adjusted HR(1)[a] (95% CI)	Adjusted HR(2)[b] (95% CI)	Adjusted HR(3)[c] (95% CI)
In hospital	94/715[d]	182/1224[d]	In hospital[e]	0.80 (0.62–1.04)	0.86 (0.66–1.12)	0.84 (0.65–1.10)	0.82 (0.63–1.07)
90 d	456.6	610.4	<90 d	**0.75 (0.61–0.92)**	**0.81 (0.66–0.99)**	**0.80 (0.65–0.98)**	**0.78 (0.64– 0.96)**
1 y	228.3	280.5	90 d–1 y	0.94 (0.74–1.18)	1.01 (0.80–1.27)	1.0 (0.79–1.26)	0.96 (0.76–1.21)
5 y	161.2	194.1	1–5 y	**0.85 (0.74–0.98)**	0.93 (0.81–1.06)	0.92 (0.80–1.05)	0.89 (0.78–1.02)

* The bolded *P* values denote statistical significance. 95% CI, 95% confidence interval; COPD, chronic obstructive pulmonary disease; HR, hazard ratio; OR, odds ratio; PY, patient year; RA, rheumatoid arthritis; SES, social economic status.
[a] Cox proportional hazard models (1) are adjusted for age and sex.
[b] Cox proportional hazard models (2) are adjusted for age, sex, SES, rural/urban, hospital size, and fracture type.
[c] Cox proportional hazard models (3) are adjusted for age, sex, SES, rural/urban, hospital size, fracture type, and comorbidities influencing risk of death: Charlson (excl. RA), cardiovascular disease, COPD, malignancy, dementia, alcoholism, diabetes, hyperlipidemia, hormonal replacement therapy, and anticoagulants.
[d] The numerator represents the number of in-hospital deaths over the total number of deaths within 5 years.
[e] In-hospital mortality was estimated using logistic regression models, and values represent unadjusted and adjusted ORs.

DISCUSSION

In this population-based study, we found that the incident RA cohort had a 28% higher risk of hip fracture than age-matched and sex-matched controls from the general population. Despite being matched on age, the RA cohort, on average, fractured their hips two years earlier than the matched general controls, indicating a premature risk of hip fractures. The incidence rate of hip fracture observed in our cohort was lower than the pooled incidence rates reported in a systematic review (4.33 [95% CI 2.26–8.27])[2] and lower than the risk ratio reported in another systematic review (risk ratio 1.28 [95% CI 1.20–1.37] vs pooled risk ratio 2.64 [95% CI 2.19–3.17]).[16] Several of the included articles in the systematic reviews were based on prevalent RA samples or in samples with differing characteristics, methods used for case ascertainment, and/or comparison/control groups.

Although the absolute risk of hip fracture increased with age for both groups, the relative risk in RA as compared with the general population decreased with increasing age and was no longer higher in individuals with RA above the age of 80 years. This is likely attributed to the low absolute risk of hip fractures at younger ages in the general population[7] leading to a higher relative risk in younger age groups with RA because of a greater risk attributable to RA. The incidence rates, however, are higher in women than men in both the general population[7] and RA cohorts.[15,38] Our findings are consistent with others,[38,39] showing a higher relative risk of hip fracture in men than in women with RA relative to the general population.

In the general population excess mortality after hip fracture is reported compared with age-matched population norms without fractures and the elevated risk persists for several years after the fracture.[9] Excess mortality has also been reported to be higher in younger age groups[9] and in men compared with women regardless of age.[7,28] Few studies have reported on mortality rates after hip fracture in RA. Studies comparing mortality in RA and control groups after hip fracture using data from Taiwan and Korean national databases reported a higher one-year mortality than control groups.[14,15] Similar to our RA cohort findings (19.2%), Gundel et al reported 90-day mortality (17.1%; 95% CI 15.7–18.5) post hip fracture in a large Danish cohort with rheumatic diseases.[40] We observed no significant differences in mortality risk post fracture between RA and general population controls who sustained a hip fracture. Although a slightly lower mortality risk within 90 days of fracture was observed in our RA sample (aHR [95% CI] 0.80 [0.65–0.98]), the clinical significance of this finding is unclear given the small magnitude of the effect, the lack of biologically plausible explanation, and the borderline statistical significance. This finding did not change when analyses were adjusted for comorbidities associated with mortality risk measured at the time of fracture, thus reducing the likelihood of a selection bias, such as a collider stratification bias, because of the sample for postfracture mortality analyses being selected based on the presence of hip fracture, but the possibility of this bias still remains.[41] Other considerations regarding differences between our results and those reported in the literature include well-recognized heterogeneity, with large geographic and ethnic variations in fracture incidence estimates and mortality post hip fracture.[42] Also of note, a comparable percentage of patients both groups received nonoperative treatment which is, in general, indicative of a poor prognosis.[43]

Because RA is a known independent risk factor for fragility fractures,[1] our findings of increased risk for hip fracture were not unexpected. The underlying reasons for increased risk for fractures with RA are likely multifaceted. Chronic inflammation and the use of glucocorticoids have known negative effects on bone mass. Other recognized risk factors for fragility fractures in RA include sex (female), older age, inactivity, lower body mass index, and health behaviors such as smoking, which is more prevalent in RA.[44] Disease-related impairments also place people with RA at a higher risk of falls, which is an important risk factor for hip fracture. People with RA have several intrinsic and extrinsic risk factors for falls.[45] A systematic review reported that the incidence of falls with RA ranged from 10% to 50% over 6 to 12 months periods and were independently associated with age, sex, and disease duration.[45] Given the public health burden of hip fracture especially with older age, fall prevention programs have traditionally targeted older adults. Insofar as adults with RA are at high risk of falling and increase risk of fragility fractures including hip fracture, prevention programs targeting RA have been advocated within the literature.[45,46]

The strengths of this study include the population-based nature of the cohort within a universal health care system, ensuring capture of all fracture events, and representativeness of the sample, including the full spectrum of RA disease severity and age groups. The use of an incident cohort of RA avoided immortal time bias when estimating fracture risk by ensuring all fracture events, including those leading to mortality, were captured. The large sample and long follow-up time allowed sufficient power and time to provide an accurate estimate of fracture rates and of mortality post fracture. Another strength was that the availability of general population controls matched on age, sex, and index RA year, which allowed comparable controls for estimating comparison of fracture risk.

Despite the strengths, some limitations are inherent to administrative data studies, including a degree of uncertainty regarding the RA diagnosis and lack of information on disease activity. We used criteria that had been validated in a subsample who participated in a RA survey, using opinion of an independent rheumatologist reviewing medical records from their treating physicians as gold standard, where the PPV was 0.82.[18] We conducted sensitivity analyses on an RA subsample who received DMARDs, and results did not substantially differ from the main analyses. Notwithstanding, the inclusion of non-RA cases in the sample would bias the results toward the null. Because the onset of RA in this cohort was identified using administrative data, a lag time between the actual symptom onset and diagnosis may exist. The possibility of initial misdiagnosis, with some patients having

another initial musculoskeletal diagnosis, cannot be excluded. Limitations relative to the hip fracture outcomes evaluated in this study include the fact that we were only able to evaluate the type of fracture and fixation, and no other information regarding the surgery and postoperative complications were available. Finally, only all-cause mortality was examined because of the relatively low number of events leading to insufficient power to accurately evaluate differences in cause-specific mortality.

The findings have important clinical implications for the management of RA. Determining the incidence of and mortality rates post hip fractures in RA relative to the general population is an initial step necessary for future initiatives targeting the prevention of hip fractures in RA. Hip fracture in this patient population likely leads to further functional limitations in a patient population already at risk of disability from their arthritis. This also has potentially important socioeconomic ramifications and subsequent consequences on health-related quality of life. Given the sparce evidence on recovery after hip fracture in RA populations, further prospective studies are warranted to determine whether the long-term recovery after a hip fracture in RA differs from the reported recovery in the general population.

In conclusion, the risk of hip fracture in our RA incident cohort was 28% greater than age-matched and sex-matched controls from the general population. Individuals with RA were more likely to fracture at an earlier age; however, after adjustment, the risk of mortality post hip fracture was comparable with the controls, except for an unexpected slightly lower 90-day mortality risk in RA relative to controls, a finding of unclear clinical significance. Although not specifically evaluated in this study, significant burden is usually seen post hip fracture with long recovery periods and commonly a lack of return to the prefracture functional level. From clinical and health policy perspectives, fall prevention interventions specific to RA populations, along with prevention programs addressing osteoporosis risk factors and facilitating early detection and management of osteoporosis, are needed to prevent hip fractures in 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 Lacaille 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.

ACKNOWLEDGMENTS

Access to data provided by the Data Stewards is subject to approval but can be requested for research projects through the Data Stewards or their designated service providers. The following data sets were used in this study: BC Cancer, Consolidation, Hospital Separations, Income Band, Lifelabs, Medical Services Plan, Pharmacare, Pharmanet, Vital Statistics-Death. You can find further information regarding these data sets by visiting the PopData project webpage at: https://my.popdata.bc.ca/project_listings/14-131/1985/01/01 - 2018/12/31. All inferences, opinions, and conclusions drawn in this publication are those of the author(s), and do not reflect the opinions or policies of the Data Steward(s).

REFERENCES

1. Kanis JA, McCloskey EV, Johansson H, et al. Development and use of FRAX® in osteoporosis. Osteoporos Int 2010;21(suppl 2):S407–S413.
2. Jin S, Hsieh E, Peng L, et al. Incidence of fractures among patients with rheumatoid arthritis: a systematic review and meta-analysis. Osteoporos Int 2018;29:1263–1275.
3. Cooper C, Cole ZA, Holroyd CR, et al. Secular trends in the incidence of hip and other osteoporotic fractures. Osteoporos Int 2011;22:1277–1288.
4. Leibson CL, Tosteson AN, Gabriel SE, et al. Mortality, disability, and nursing home use for persons with and without hip fracture: a population-based study. J Am Geriatr Soc 2002;50:1644–1650.
5. Hauser B, Riches PL, Wilson JF, et al. Prevalence and clinical prediction of osteoporosis in a contemporary cohort of patients with rheumatoid arthritis. Rheumatology (Oxford) 2014;53:1759–1766.
6. Wu A-M, Bisignano C, James SL, et al; GBD 2019 Fracture Collaborators. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet Healthy Longev 2021;2(9):e580–e592.
7. Sing CW, Lin TC, Bartholomew S, et al. Global epidemiology of hip fractures: secular trends in incidence rate, post-fracture treatment, and all-cause mortality. J Bone Miner Res 2023;38:1064–1075.
8. Downey C, Kelly M, Quinlan JF. Changing trends in the mortality rate at 1-year post hip fracture - a systematic review. World J Orthop 2019;10:166–175.
9. Abrahamsen B, van Staa T, Ariely R, et al. Excess mortality following hip fracture: a systematic epidemiological review. Osteoporos Int 2009;20:1633–1650.
10. Omsland TK, Emaus N, Tell GS, et al. Mortality following the first hip fracture in Norwegian women and men (1999–2008). A NOREPOS study. Bone 2014;63:81–86.
11. Magaziner J, Hawkes W, Hebel R, et al. Recovery from hip fracture in eight areas of function. J Gerontol A Biol Sci Med Sci 2000;55:M498–M507.
12. Naglie G, Tansey C, Kirkland JL, et al. Interdisciplinary inpatient care for elderly people with hip fracture: a randomized control trial. CMAJ 2002;167:25–32.
13. Mazzucchelli R, Pérez Fernandez E, Crespí-Villarías N, et al. Trends in hip fracture in patients with rheumatoid arthritis: results from the Spanish National Inpatient Registry over a 17-year period (1999–2015). TREND-AR study. RMD Open 2018;4(1):e000671.
14. Lin YC, Li YH, Chang CH, et al. Rheumatoid arthritis patients with hip fracture: a nationwide study. Osteoporos Int 2015;26:811–817.
15. Kwon H-Y, Kim H-H, Sung Y-K, et al. Incidence and mortality of osteoporotic fracture in rheumatoid arthritis in South Korea using nationwide claims data. J Bone Metab 2019;26:97–104.
16. Xue AL, Wu S-Y, Jiang L, et al. Bone fracture risk in patients with rheumatoid arthritis: a meta-analysis. Medicine (Baltimore) 2017;96:e6983.
17. Yazdani K, Xie H, Avina-Zubieta JA, et al. Has the excess risk of acute myocardial infarction in rheumatoid arthritis relative to the general population declined? A population study of trends over time. Semin Arthritis Rheum 2021;51:442–449.

18. Chang J, Rogers P, Lacaille D. Can American College of Rheumatology criteria for Rheumatoid Arthritis be assessed using self-report? Comparison of self-reported data with chart review. Arthritis Rheum 2011;63(suppl):S49.
19. Canada S. British Columbia (Code 59) and Canada (Code 01) (table). Census Profile. Census. Statistics Canada Catalogue ed. Ottawa, ON: Statistics Canada; 2018.
20. Canadian Institute for Health Information. Discharge Abstract Database (Hospital Separations): Population Data BC 2017. Data Extract. MOH. Accessed December 11, 2024. http://www.popdata.bc.ca/data
21. British Columbia Ministry of Health. Medical Services Plan (MSP) Payment Information File. V2: Population Data BC 2016. Data Extract. MOH. Accessed December 11, 2024. https://www.popdata.bc.ca/data
22. BC Vital Statistics Agency. Vital statistics deaths: population data BC 2017. Data Extract BC Vital Statistics Agency. Accessed December 11, 2024. http://www.popdata.bc.ca/data
23. British Columbia Ministry of Health. PharmaNet. V2 [Data Extract]. Data Stewardship Committee; 2016. Data Extract. MOH. Accessed December 11, 2024. https://www.popdata.bc.ca/data
24. Lix LM, Azimaee M, Osman BA, et al. Osteoporosis-related fracture case definitions for population-based administrative data. BMC Public Health 2012;12:301.
25. Sobolev B, Guy P, Sheehan KJ, et al; Canadian Collaborative Study on Hip Fractures. Time trends in hospital stay after hip fracture in Canada, 2004–2012: database study. Arch Osteoporos 2016;11:13.
26. Hudson M, Avina-Zubieta A, Lacaille D, et al. The validity of administrative data to identify hip fractures is high - a systematic review. J Clin Epidemiol 2013;66:278–285.
27. Pulcins I, Wan E. In-hospital hip fractures in Canada: using information to improve patient safety. Healthc Q 2004;7:25–27.
28. Haentjens P, Magaziner J, Colón-Emeric CS, et al. Meta-analysis: excess mortality after hip fracture among older women and men. Ann Intern Med 2010;152:380–390.
29. Sheehan KJ, Sobolev B, Guy P, et al; Canadian Collaborative Study on Hip Fractures. Constructing an episode of care from acute hospitalization records for studying effects of timing of hip fracture surgery. J Orthop Res 2016;34:197–204.
30. Yurkovich M, Avina-Zubieta JA, Thomas J, et al. A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol 2015;68:3–14.
31. Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993;46:1075–1079.
32. Romano PS, Roost LL, Jollis JG. Further evidence concerning the use of a clinical comorbidity index with ICD-9-CM administrative data. J Clin Epidemiol 1993;46:1085–1090.
33. Marozoff S, Mai A, Dehghan N, et al. Increased risk of venous thromboembolism in patients with granulomatosis with polyangiitis: a population-based study. PLoS One 2022;17:e0270142.
34. Bursac Z, Gauss CH, Williams DK, et al. Purposeful selection of variables in logistic regression. Source Code Biol Med 2008;3:17.
35. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. J Am Stat Assoc 1999;94:496–509.
36. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol 2009;170:244–256.
37. Gray RJ. A class of K-Sample Tests for comparing the cumulative incidence of a competing risk. Ann Stat 1988;16:1141–1154.
38. Kim SY, Schneeweiss S, Liu J, et al. Risk of osteoporotic fracture in a large population-based cohort of patients with rheumatoid arthritis. Arthritis Res Ther 2010;12:R154.
39. Yamamoto Y, Turkiewicz A, Wingstrand H, et al. Fragility fractures in patients with rheumatoid arthritis and osteoarthritis compared with the general population. J Rheumatol 2015;42:2055–2058.
40. Gundel O, Thygesen LC, Gögenur I, et al. Postoperative mortality after a hip fracture over a 15-year period in Denmark: a national register study. Acta Orthop 2020;91:58–62.
41. Banack HR, Kaufman JS. The obesity paradox: understanding the effect of obesity on mortality among individuals with cardiovascular disease. Prev Med 2014;62:96–102.
42. Kanis JA, Odén A, McCloskey EV, et al. A systematic review of hip fracture incidence and probability of fracture worldwide. Osteoporos Int 2012;23:2239–2256.
43. Loggers SAI, Van Lieshout EMM, Joosse P, et al. Prognosis of nonoperative treatment in elderly patients with a hip fracture: a systematic review and meta-analysis. Injury 2020;51:2407–2413.
44. Baker R, Narla R, Baker JF, et al. Risk factors for osteoporosis and fractures in rheumatoid arthritis. Best Pract Res Clin Rheumatol 2022;36:101773.
45. Brenton-Rule A, Dalbeth N, Bassett S, et al. The incidence and risk factors for falls in adults with rheumatoid arthritis: a systematic review. Semin Arthritis Rheum 2015;44:389–398.
46. Stanmore EK, Oldham J, Skelton DA, et al. Fall incidence and outcomes of falls in a prospective study of adults with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2013;65(5):737–744.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 614–622
DOI 10.1002/acr.25480

American College of Rheumatology
Empowering Rheumatology Professionals

Association of Changes in Hand Pain With BMI, Employment, and Mental Well-Being Over Four Years in Patients With Hand Osteoarthritis

Coen van der Meulen,[1] Lotte A. van de Stadt,[2] Saskia J. Buck,[1] Frits R. Rosendaal,[1] Sietse E. S. Terpstra,[1] and Margreet Kloppenburg[1]

Objective. We aimed to characterize patients with hand osteoarthritis (OA) with deteriorating or improving hand pain and to investigate patients achieving good clinical outcome after four years.

Methods. We used four-year annual Australian/Canadian Hand Osteoarthritis Index (AUSCAN) pain subscale (range 0–20) measurements from the Hand OSTeoArthritis in Secondary Care cohort (patients with hand OA). Pain changes were categorized as deterioration, stable, and improvement using the Minimal Clinical Important Improvement. Good clinical outcome was categorized using the Patient Acceptable Symptom State (PASS). Associations between baseline characteristics (patient and disease characteristics, coping styles, and illness perceptions) and outcomes were investigated using multinomial or binary logistic regression, adjusted for baseline pain, age, sex, and body mass index (BMI).

Results. A total of 356 patients (83% female, mean age 60.6 years, mean AUSCAN score 9.1) were analyzed. Pain improved for 38% of patients, deteriorated for 30% of patients, and remained stable for 32% of patients over four years. Four-year pain development followed annual trends. At baseline, 44% of patients reached PASS, and 49% of patients reached PASS at follow-up. Higher BMI, coping through comforting cognitions, and illness comprehension were positively associated with pain deterioration. Higher AUSCAN function score, mental well-being, and illness consequences were negatively associated with pain improvement. Employment (positive) and emotional representations (negative) were associated with both improvement and deterioration. Higher baseline AUSCAN function, tender joint count, and symptoms attributed to hand OA were associated negatively with PASS after four years.

Conclusion. The pain course of patients with hand OA is variable, not inevitably worsening, and various factors may play a role. Whether modification of these risk factors can influence pain outcomes requires further investigation.

INTRODUCTION

Osteoarthritis (OA) is a chronic disease that progresses over the course of multiple years. Hand OA is a prevalent OA subtype, resulting in structural damage and symptoms including disability, loss of quality of life, and pain in the hand.[1,2] Different processes are thought to underly this hand pain, such as nociceptive pain (both mechanical and inflammatory in origin) and nociplastic pain (due to sensitization).[3–5] Inflammatory pain can arise from local processes, for example, synovitis,[6] or from systemic processes, such as obesity and the accompanying adipokines.[7] Mechanic pain can arise through structural damage to the joint and mechanical loading developed, for example, during intense manual labor.[8,9] Mental factors are also thought to contribute to pain in OA. These include coping styles and illness perceptions.[8,10,11] Because of its multifactorial nature and the plethora of underlying mechanisms, treating pain in hand OA is challenging.

Little is known about the course of hand OA pain over time and what determines this course. Given the chronic nature of the disease and the known gradual increase in structural damage, one would expect the pain to increase over time. However,

Supported by the Dutch Arthritis Society.
[1]Coen van der Meulen, MD, Saskia J. Buck, MSc, Frits R. Rosendaal, MD, PhD, Sietse E. S. Terpstra, MD, Margreet Kloppenburg, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands; [2]Lotte A. van de Stadt, MD, PhD: Amsterdam Rheumatology and Immunology Center, Reade, Amsterdam, The Netherlands.
Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25480).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25480.
Address correspondence via email to Coen van der Meulen, MD, at c.van_der_meulen@lumc.nl.
Submitted for publication October 10, 2023; accepted in revised form December 4, 2024.

SIGNIFICANCE & INNOVATIONS

- Similar numbers of patients show an increase, a decrease, or a stable course of pain over four years.
- The number of patients at an acceptable level of pain increased from 44% to 49% over four years.
- Changes in pain were associated with body mass index, employment status, mental well-being, illness perceptions, and coping styles.
- These factors may be used for patient stratification, both in clinical and research settings.

studies on the development of pain both over the short term (2 years) and the long term (6 or 10 years) found that pain on the group level largely remains stable.[12–14] Similar results were seen over four years in a previous study by our group.[15] However, individual patients may experience a deterioration or an improvement in pain.[12] These patients can be categorized using the minimal clinically important improvement (MCII) score, a measure used to categorize clinically meaningful improvements, enabling the investigation of changes on the patient level.[16] Little is known about what influences the development of pain. Change in pain has been associated with change in synovitis measured on magnetic resonance imaging (MRI), but not with radiographic signs.[17] It is currently unknown whether other mechanisms and risk factors that contribute to the occurrence of pain also influence the course of pain.

Perhaps even more clinically relevant for the patient is the acceptability of a given level of pain. The limit of acceptability for a symptom can be defined with the Patient Acceptable Symptom State (PASS). The PASS describes the highest level of symptoms at which patients regard the symptom as acceptable, should it remain at that level for the rest of their life.[16] Little is known regarding determinants of reaching a PASS for pain in hand OA, with a previous study showing that patients with worse pain and function scores at baseline, as well as more painful joints at baseline, were less likely to reach PASS after six years.[12]

Therefore, this study aimed to evaluate the change in hand pain on the mid-term for individual patients, and, after four years, to characterize the patients with improving or deteriorating pain and to investigate which patients are likely to achieve a good clinical outcome. We aimed to identify potential modifiable risk factors, to support the move toward personalized medicine, and to enable optimal patient inclusion in clinical trials.

PATIENTS AND METHODS

Study design. The data were derived from the Hand OSTeoArthritis in Secondary Care (HOSTAS) cohort. The HOSTAS is an observational cohort study of consecutively referred patients with primary hand OA, collected from the Leiden University Medical Center rheumatology outpatient clinic between June 2009 and October 2015. The HOSTAS included patients who had a clinical diagnosis of hand OA, determined by their treating rheumatologist. Exclusion criteria included any pathologic conditions that could otherwise explain the symptoms of the hand (eg, carpal tunnel syndrome, strain, fibromyalgia, and other rheumatic musculoskeletal diseases) and secondary OA (eg, due to inflammatory joint diseases such as rheumatoid arthritis or psoriatic arthritis, bone diseases such as osteitis deformans and osteochondritis, fractures, metabolic diseases such as hemochromatosis, bone dysplasia, endocrine diseases such as acromegaly, major congenital or developmental diseases, and major local diseases such as hypermobility or gout). Finally, patients with a language barrier or psychologic limitations precluding participation or informed consent were excluded. Patients answered questionnaires yearly and underwent physical examinations biannually for four years. Full details on the cohort have been published previously.[18]

The HOSTAS cohort consists of 538 patients. Only patients with Australian/Canadian Hand Osteoarthritis Index (AUSCAN) pain measurements at both baseline and year 4 (required for the main outcome) were included in the analysis (n = 356). The HOSTAS study was approved by the medical ethics committee at the Leiden University Medical Center and conducted in accordance with Good Clinical Practice guidelines and the Declaration of Helsinki. All patients provided written informed consent.

Outcome. The primary outcome for this study was the validated AUSCAN pain score. The AUSCAN pain score is calculated by summing five individual component questions, each worth 0 to 4 points, for a total score of 0 to 20, with higher scores indicating more pain, collected through a questionnaire.[19] The AUSCAN questionnaire further contains a function domain (nine questions, total score 0–36).

Covariates. Additional validated questionnaires collected included the Hospital Anxiety and Depression Scale (HADS) for signs of anxiety and depression (seven questions for each domain, scored 0–3 for 0–21 domain scores).[20] Illness perceptions associated with hand OA were investigated using the Illness Perception Questionnaire (IPQ), with questions stating that they concerned hand OA.[21] This questionnaire measures eight domains of illness perceptions and attributions, with higher scores indicating stronger beliefs in the investigated concept. A detailed explanation is attached in the Supplementary Methods. Coping strategies were investigated using the Coping with Rheumatic Stressors (CORS) questionnaire, which investigates eight different coping styles and was developed for use in rheumatic musculoskeletal diseases.[22] Higher scores indicate more use of a particular coping style. Details can be found in the Supplementary File. Demographic information including age, sex, marital status (categorized into married and/or living together or not),

working status (categorized into currently working or not (meaning currently not employed, work disabled or with sick leave), excluding pensioners and patients replying "other" from the analysis), and education level (categorized into low education level and other [mid and high education level]) were collected through a questionnaire. Furthermore, the time of first symptoms was collected from this questionnaire and used to calculate symptom duration. Height and weight were measured and used to calculate body mass index (BMI). Information on comorbidities was collected using the modified Charlson Index.[23] Given the distribution of the information on comorbidities, this was dichotomized to comorbidities or no comorbidities for the analyses. Finally, information on use of analgesics (including paracetamol, nonsteroidal anti-inflammatory drugs, opioids, or other types of analgesics) was collected by the questionnaire.

Radiographic signs (Kellgren-Lawrence sum score over 30 joints[24] and presence of erosive disease in interphalangeal joints according to Verbruggen-Veys [defined as ≥1 interphalangeal joint in the E or R phase[25]]) were investigated from hand radiographs made at baseline. T2 MRI images without contrast, made using a 1.5T MRI scanner, were scored while masked for patient characteristics using the Hand Osteoarthritis Magnetic Resonance Imaging Scoring System for synovitis and effusion, with the distal interphalangeal (DIP) and proximal interphalangeal (PIP) joints scored on a scale of 0 to 3 for the right hand.[26] Reliability of scoring was excellent, with an intraclass correlation of 0.93, based on 25 scans scored twice in a random order. Physical examination was performed to determine the tender joint count and establish fulfillment of the American College of Rheumatology (ACR) criteria.[27]

Annual data from the AUSCAN pain score from baseline up to and including year 4 were used. Baseline data were used for all other variables.

AUSCAN pain scores were regarded as missing when more than one component question was missing, or two in case of the function scores. IPQ scores were regarded as missing in case of one or two missing components, depending on the domain. HADS and CORS scores were regarded as missing if any component question was missing. Missing data were less than 5% for most variables. The few variables with more than 5% missing are explained by the addition of those specific measurements after the start of data collection. Patients who entered before the inclusion of these measurements have missing values for these variables. These missing values are considered missing completely at random. Another variable with more than 5% missing data was "currently working," in which retired patients were excluded. These patients were not part of the group of interest. Missing data were not imputed.

Statistical analysis. Baseline characteristics were described using means with SDs or medians with interquartile ranges for continuous measures, as appropriate, and as numbers with percentages for categorical variables. Change scores in AUSCAN pain from baseline to year 4 were calculated and used to classify patients as having stable, deteriorated, or improved pain after four years of participation in the study, based on the MCII, previously established[28] at 1.6. Based on this classification, an increase of >1.6 was classified as a deterioration of the pain status of the patient, a decrease less than −1.6 as an improvement, and changes between −1.6 and 1.6 were classified as stable. Good clinical outcome was categorized according to the PASS of 8.2, with scores lower than this cutoff counting as having attained PASS.[28]

Annual changes in pain in the three change groups (stable, improvement, and deterioration) were visualized using heatmaps. Differences between patients experiencing a deterioration or an improvement in pain from baseline to year 4 and the stable group were investigated using multinomial logistic regression analysis, with the change categories stable, improvement, or deterioration as the dependent, and no change in pain (the stable group) as the index. Baseline variables hypothesized to influence pain development in hand OA were tested and used as independent variables. Separate models were run for independent variables, with each model adjusted for baseline pain, age, sex, and BMI. Associations between the independent variables and change in pain from baseline to year 4 were determined using the adjusted odds ratios (ORs) obtained from these multinomial logistic regression models. Baseline values for all independent variables were used. Analyses for variables showing an effect in these analyses were repeated after stratification for the presence of comorbidities and use of analgesics at baseline.

Patients reaching the PASS at year 4 were investigated using binary logistic regression, with patients not reaching the PASS as the index group. The same variables hypothesized to influence pain in hand OA were tested in separate models, adjusted for baseline age, sex, baseline BMI, and baseline pain. All analyses were performed using RStudio running R version 4.0.3.

RESULTS

Patient characteristics. An overview of patient characteristics at entry is shown in Table 1. Of 356 patients, there were 296 female participants (83%), and the mean age was 60.6 (SD ±8.2) years (range 39.6–86.3 years). The ACR criteria were fulfilled by 326 of the cohort (92%), and the mean AUSCAN pain score at baseline was 9.1 (SD ±4.3). There were no major differences between the included and excluded patients for the longitudinal analysis groups (Supplementary Table S1).

Annual change in AUSCAN pain between visits. Changes over four years were used to categorize patients into improvement, deterioration, and stable groups (Supplementary Table S2). Over four years, AUSCAN pain improved in 137 patients (38%, mean baseline pain 10.8 [SD ±3.9], mean

Table 1. Patient characteristics at baseline*

Characteristics	Patients (N = 356)
Patient characteristics	
Female sex, n (%)	296 (83)
Age, mean (SD), yr	60.6 (8.2)
BMI, mean (SD)	26.8 (4.5)
Married or living together, n (%)	291 (82)
Low education level, n (%)	80 (23)
Currently working, n (%)[a]	163 (75)
Number of comorbidities (range 0–18), median (IQR)	0 (0–1)
Presence of any comorbidities, n (%)	146 (42)
Current use of analgesics, mean (SD)	232 (65)
Disease characteristics	
ACR criteria fulfilled, n (%)	326 (92)
Symptom duration, median (IQR), yr	5.6 (2.0–12.6)
Erosive disease, n (%)	106 (30)
KL sum score (range 0–120), median (IQR)	17 (9–31)
Synovitis on MRI (range 0–24), median (IQR)	0 (0–1)
If any synovitis present (range 0–24)	2 (1–3)
Tender joint count (range 0–30), median (IQR)	3 (1–6)
Patient reported outcome measures	
AUSCAN	
Pain (range 0–20), mean (SD)	9.1 (4.3)
PASS at baseline, n (%)	155 (44)
Function (range 0–36), mean (SD)	14.9 (8.4)
HADS, median (IQR)	
Depression (range 0–21)	2 (1–5)
Anxiety (range 0–21)	4 (2–7)

* Percentage of missing data was lower than 5%, unless indicated otherwise. For the currently working category, n = 216; for synovitis, n = 207; and for HADS, n = 254. ACR, American College of Rheumatology; AUSCAN, Australian/Canadian Osteoarthritis Hand Index; BMI, body mass index; HADS, Hospital Anxiety and Depression Scale; IQR, interquartile range; KL, Kellgren-Lawrence; MRI, magnetic resonance imaging; PASS, Patient Acceptable Symptom State.
[a] Excluding pensioners.

change score −4.8 [SD ±2.8]), was stable in 113 patients (32%, mean baseline pain 8.6 [SD ±4.3], mean change score −0.1 [SD ±0.8]), and deteriorated in 106 patients (30%, mean baseline pain 7.3 [SD ±3.9], mean change score 3.8 [SD ±1.9]). Annual changes also reflected the course of the changes over four years, as can be seen in Figure 1. These annual changes in pain ranged from −12 to +11. Comparing the annual changes among the groups, patients with improvement over four years showed improvement (decreased pain) on annual intervals (green bars) more frequently than patients in the stable and deterioration groups. Similarly, patients in the deterioration group showed more annual intervals with deterioration (increased pain, red bars). The stable pain group showed the most heterogeneity in annual intervals of the three groups.

Associations with a deterioration or an improvement in pain. BMI at baseline was positively associated with deterioration in pain (OR 1.08), with an increase of 1 multiplying the odds of experiencing a deterioration in pain with 1.08, or an increase of 5 multiplying the odds by 1.40. Age and sex were not associated with changes in pain over four years. Patients who were currently employed had a higher chance to report either deteriorated or improved pain after four years than to report a stable level of pain, whereas those unemployed or on sick leave were more likely to report stable levels of pain (Table 2).

Only a few of the patient-reported outcome measures showed an association with a deterioration in pain. Use of the coping style comforting cognitions and the IPQ domain "illness coherence" (with higher scores indicating less understanding of the disease) were positively associated with a deterioration in pain over four years, whereas the IPQ domain "emotional representations" (with higher scores indicating more negative emotions attributed to hand OA) was negatively associated with a deterioration in pain. Baseline AUSCAN function (higher scores equals worse function), HADS depression and anxiety scores (higher scores equal more signs of depression and anxiety), and the IPQ domains "emotional representations" and "consequences" (measuring consequences attributed to hand OA) were all negatively associated with an improvement in pain after four years.

No associations were found for disease characteristics, including erosive disease, synovitis, or radiographic signs with changes in pain (Table 2). Models that showed effects on change in pain were stratified for the presence or absence of comorbidity and use or non-use of analgesics. The stratified analyses showed similar results as the unstratified analyses (Supplementary Tables S3–S9).

Associations with good clinical outcome. At baseline, 155 patients (44%) were at PASS. At year 4, 176 patients (49%) were at PASS. Of those at PASS at baseline, 112 patients (72%) were still at PASS at year 4. A total of 43 patients lost PASS, and 64 patients reached PASS. Being at PASS at baseline was strongly associated with being at PASS at year 4, but the effect attenuated with adjustment (crude OR 5.6 [95% confidence interval 3.5–8.9]; OR adjusted for age, sex, and BMI 6.1 [95% confidence interval 3.8–9.9]; OR adjusted for baseline pain, age, sex, and BMI 1.21 [95% confidence interval 0.53–2.77]). Patients with worse hand function (AUSCAN function score) and higher tender joint count at baseline were less likely to reach a good clinical outcome at year 4. The identity scale of the IPQ, indicating how many symptoms are considered related to the OA by the patient, was also negatively associated with reaching good clinical outcome at year 4. No further associations were seen (Table 3).

DISCUSSION

This study aimed to investigate pain development and good clinical pain outcomes in patients with hand OA. We observed 356 patients with hand OA over four years and found that 137 patients experienced an improvement in pain, 106 patients experienced a deterioration, and 113 patients experienced a

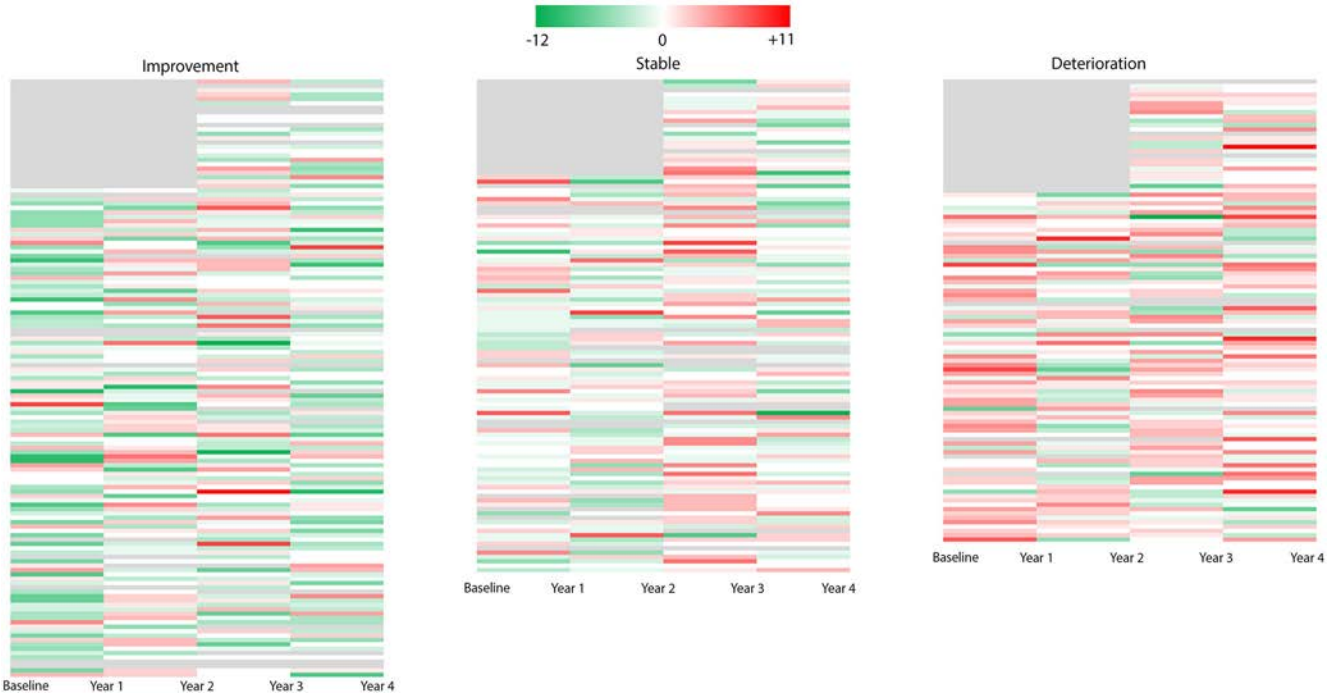


Figure 1. Heatmaps for annual change in Australian/Canadian Hand Osteoarthritis Index (AUSCAN) pain between visits. Change in AUSCAN pain between visits for individual patients. In each panel, the first column indicates change from baseline to year 1, the second is year 1 to year 2, the third is year 2 to year 3, and the final column is year 3 to year 4. White indicates a change of 0, green indicates improvement in pain, and red indicates deterioration. Gray indicates a missing value. Darker color means a larger change (range from −12 to +11). Heatmaps show data for patients that had an improvement in pain (n = 137; left), patients with stable pain (n = 113; middle), and patients with a deterioration in pain (n = 106; right). Categorization on change from baseline to year 4, with the Minimal Clinical Important Improvement of 1.6 as the cutoff.

stable level of pain. The changes over four years were consistent with annual changes. An improvement in pain after four years (38% of patients) was seen most in patients who, at baseline, had a better hand function, fewer mental problems, a paid job, and attributed fewer negative emotions and consequences to the disease than other patients. A deterioration in pain over four years was seen in 30% of patients, and this was most prevalent in patients with a higher BMI and a job at baseline, who used comforting cognitions as a coping strategy and who perceived they understood the disease better. A good clinical outcome, defined as being at PASS after four years, was seen in 49% of patients, a slight increase from 44% at baseline. Being at PASS at baseline was strongly associated with being at PASS at year 4. Furthermore, patients with better hand function and fewer painful joints on palpation and who attributed less symptoms to their hand OA at baseline were more likely to be at PASS at year 4.

Previously, we found that pain remained stable on a group level in patients with hand OA over four years.[15] This stable group level may mask a mixture of patients experiencing a deterioration or an improvement. This was confirmed by our findings reported here. Part of the change in pain over time can be explained by regression to the mean. Investigation of the annual changes showed that changes over four years were consistent with annual changes, as well. Participants experiencing a deterioration over four years also experienced a deterioration in pain per year more often than patients in the stable and improvement groups. This makes it unlikely that the changes seen in this study are solely due to regression to the mean or chance. A previous study with six years' follow-up reported more participants experiencing deterioration in pain (40%) than improvement (26%).[12] There may be a number of explanations: the difference in average baseline pain between the studies (6.7 in the previous study vs 9.1 in the current study), differences in type of patients with OA studied (hand OA vs polyarticular familiar hand OA), or the difference in follow-up duration (four vs six years).

It should be noted that changes in pain were based on yearly measurements in this study, which may not adequately capture fluctuations in pain between these time points. An alternative approach might be a pain diary, which provides more detailed data. However, a diary is very time consuming for participants. It is also likely to bias the results due to response shift.[29] Filling in a pain diary may also influence pain awareness, leading to further bias. This makes it difficult to determine the optimal interval duration for pain questionnaires.

The changes over time found in this study were associated with various factors. BMI was associated with a deterioration in pain after four years. The positive association between BMI and pain in hand OA has previously been reported cross-sectionally

Table 2. Associations of baseline characteristics with clinically important deterioration or improvement in pain*

	Adjusted odds ratio (95% confidence interval)	
	Deterioration (n = 106)	Improvement (n = 137)
Patient characteristics at baseline		
Female sex[a]	1.02 (0.49–2.15)	0.80 (0.39–1.63)
Age, yr[b]	0.98 (0.94–1.01)	0.99 (0.96–1.02)
BMI[c]	1.08 (1.01–1.14)	0.99 (0.93–1.05)
Married or living together	1.79 (0.82–3.88)	1.01 (0.53–1.94)
Low education level	0.99 (0.50–1.96)	1.05 (0.54–2.03)
Currently working[d]	3.35 (1.39–8.11)	4.44 (1.83–10.7)
Presence of any comorbidities	1.03 (0.57–1.85)	0.98 (0.57–1.70)
Current use of analgesics	0.81 (0.44–1.50)	0.96 (0.53–1.75)
Disease characteristics at baseline		
Erosive disease present	1.19 (0.64–2.22)	0.87 (0.48–1.57)
KL sum score (range 0–120)	1.01 (0.99–1.03)	1.00 (0.98–1.02)
Symptom duration, yr	1.02 (0.99–1.06)	0.97 (0.94–1.01)
Tender joint count (range 0–30)	1.06 (0.99–1.13)	0.96 (0.90–1.03)
Synovitis on MRI (range 0–24)	0.97 (0.78–1.21)	0.90 (0.73–1.13)
Patient reported outcome measures at baseline		
AUSCAN function (range 0–36)	1.01 (0.95–1.06)	0.95 (0.91–1.00)
HADS		
Depression (range 0–21)	0.92 (0.82–1.03)	0.89 (0.80–0.98)
Anxiety (range 0–21)	0.92 (0.82–1.03)	0.91 (0.82–1.00)
CORS		
Pain		
Comforting cognitions (9–36)	1.10 (1.02–1.19)	1.04 (0.97–1.11)
Decreasing activity (8–32)	0.99 (0.91–1.07)	0.95 (0.88–1.02)
Diverting attention (8–32)	1.04 (0.97–1.12)	1.05 (0.98–1.12)
Limitations		
Optimism (5–20)	1.07 (0.95–1.20)	1.03 (0.93–1.14)
Pacing (10–40)	1.02 (0.97–1.08)	0.96 (0.91–1.02)
Creative solutions (8–32)	1.04 (0.97–1.11)	0.99 (0.92–1.05)
Dependency		
Accepting (6–24)	1.00 (0.92–1.09)	1.00 (0.92–1.08)
Consideration (7–28)	1.02 (0.93–1.12)	0.99 (0.91–1.07)
IPQ		
Identity (0–14)	1.03 (0.86–1.24)	0.91 (0.78–1.06)
Timeline (chronic) (6–30)	1.02 (0.92–1.12)	1.00 (0.92–1.09)
Consequences (6–30)	0.94 (0.86–1.03)	0.91 (0.84–0.98)
Personal control (6–30)	0.97 (0.89–1.07)	1.04 (0.95–1.13)
Treatment control (5–25)	1.03 (0.91–1.17)	0.97 (0.87–1.09)
Illness coherence (5–25)	1.11 (1.01–1.22)	1.07 (0.99–1.17)
Timeline cyclical (4–20)	0.99 (0.90–1.10)	0.98 (0.90–1.08)
Emotional representations (6–30)	0.92 (0.86–1.00)	0.93 (0.88–1.00)

* Associations with changes in AUSCAN pain, defined by Minimal Clinical Important Improvement, adjusted for baseline AUSCAN pain, age, sex, and BMI. N = 356. Stable group as index. AUSCAN, Australian/Canadian Osteoarthritis Hand Index; BMI, body mass index; CORS, Coping with Rheumatic Stressors; HADS, Hospital Anxiety and Depression Scale: IPQ, Illness Perception Questionnaire; KL, Kellgren-Lawrence; MRI, magnetic resonance imaging.
[a] Adjusted for baseline pain, age, and BMI.
[b] Adjusted for baseline pain, sex, and BMI.
[c] Adjusted for baseline pain, age, and sex.
[d] Compared to currently not working, excluding retirees (n = 161).

and is thought to stem from the systemic inflammation caused by adipokines.[7,8] Our study contributes an association between BMI and clinically relevant changes in pain over time, which means that pain modulation by BMI may be a continuous process. Currently working was associated with both deterioration and improvement of pain. Having paid work may be a proxy for being in a more active phase of life with more varying demands. This could translate to changing pain scores, reported in response to changes in the strain placed on the hands (eg, switching from a manual to a desk job). This hypothesis requires further validation.

Having less signs of anxiety or depression (measured with the HADS) was associated with an improvement in pain, highlighting the previously described effect of overall mental well-being on pain.[8] Interestingly, better well-being at baseline was associated with an improvement in pain, whereas worse well-being showed no association with a deterioration in pain. Mental well-being is

Table 3. Associations with good clinical outcome*

	Adjusted odds ratio (95% confidence interval), PASS (n = 176)
Patient characteristics at baseline	
Female sex[a]	1.14 (0.59–2.21)
Age, yr[b]	1.00 (0.97–1.03)
BMI[c]	0.96 (0.91–1.02)
Married or living together	0.66 (0.35–1.26)
Low education level	1.09 (0.60–1.99)
Currently working[d]	1.42 (0.68–3.00)
Presence of any comorbidities	0.73 (0.44–1.22)
Current use of analgesics	0.97 (0.57–1.67)
Disease characteristics at baseline	
Erosive disease present	0.85 (0.49–1.47)
KL sum score (range 0–120)	1.00 (0.98–1.01)
Symptom duration, yr	0.98 (0.95–1.01)
Tender joint count (range 0–30)	0.91 (0.84–0.97)
Synovitis on MRI (range 0–24)	0.92 (0.75–1.12)
Patient reported outcome measures at baseline	
AUSCAN function (range 0–36)	0.95 (0.91–0.99)
HADS	
Depression (range 0–21)	0.93 (0.84–1.03)
Anxiety (range 0–21)	0.97 (0.88–1.06)
CORS	
Pain	
Comforting cognitions (9–36)	0.96 (0.90–1.03)
Decreasing activity (8–32)	0.96 (0.90–1.03)
Diverting attention (8–32)	0.98 (0.93–1.05)
Limitations	
Optimism (5–20)	0.98 (0.89–1.08)
Pacing (10–40)	0.99 (0.94–1.04)
Creative solutions (8–32)	0.98 (0.93–1.04)
Dependency	
Accepting (6–24)	1.02 (0.95–1.10)
Consideration (7–28)	0.98 (0.90–1.06)
IPQ	
Identity (0–14)	0.81 (0.69–0.95)
Timeline (chronic) (6–30)	0.96 (0.89–1.04)
Consequences (6–30)	0.96 (0.89–1.03)
Personal control (6–30)	1.01 (0.94–1.10)
Treatment control (5–25)	1.00 (0.90–1.11)
Illness coherence (5–25)	0.97 (0.90–1.05)
Timeline cyclical (4–20)	1.00 (0.92–1.09)
Emotional representations (6–30)	1.03 (0.97–1.09)

* Associations with reaching PASS at year 4, adjusted for baseline AUSCAN pain, age, sex, and BMI. N = 356. Group not reaching PASS as index. AUSCAN, Australian/Canadian Osteoarthritis Hand Index; BMI, body mass index; CORS, Coping with Rheumatic Stressors; HADS, Hospital Anxiety and Depression Scale; IPQ, Illness Perception Questionnaire; KL, Kellgren-Lawrence; MRI, magnetic resonance imaging; PASS, Patient Acceptable Symptom State.
[a] Adjusted for baseline pain, age, and BMI.
[b] Adjusted for baseline pain, sex, and BMI.
[c] Adjusted for baseline pain, age, and sex.
[d] Compared to currently not working, excluding retirees (n = 161).

reinforced as a potential therapeutic target for pain in hand OA. Whether treatment of the mental well-being of patients leads to improvement in their pain outcomes requires further investigation.

Better hand function at baseline was associated with improvement in pain. It has previously been shown that functional limitations at baseline are associated with poor pain outcomes.[12] As such, an association between better function at baseline and an improvement in pain was expected. Mechanistically, we expect that a change in function follows a change in pain. However, patients reporting worse function may feel more limited by their hand symptoms. They may then report those symptoms, including pain, as more severe due to increased attention on those symptoms. This could lead to a negative spiral and changes in reported pain. Therapy supporting hand function might relieve pain and should be investigated further.

Previous studies indicated that illness perceptions and coping styles may be targets for interventions to improve hand function.[30,31] Additional work showed that developing more negative illness perceptions was associated with a worsening of functional hand OA outcomes over six years.[32] We add to this finding that perceiving less consequences of hand OA is associated with an improvement in pain and that understanding the disease better was associated with a deterioration in pain, indicating illness perceptions can also have effects on pain. Illness perceptions may be a target to improve pain and function. The positive association of coping using comforting cognitions with a deterioration in pain reinforces that different coping styles may also influence pain development. Increasing the patient's resilience to pain through education may therefore be of value in treating pain in hand OA. It should be noted that perceiving fewer negative emotions due to the hand OA was associated with both deterioration and improvement in pain. Illness perceptions can potentially have different effects in different patients, possibly being dependent on other patient beliefs and personality traits.

The presence of comorbidities and use of analgesics were not added to the models as covariates because of the size of the confidence intervals obtained when attempting to do so. This could be explained by the small strata underlying these analyses, as the categorical variables yielded strata of <10 participants when combined, before adding continuous variables (Supplementary Tables S10 and S11). We employed stratification instead. Some variables showed slightly different associations over the strata, but no major differences or changes in the direction of the association were seen. The resulting confidence intervals were wider, which can be explained by the smaller number of participants per stratum of the analyses. The low number of participants per stratum precludes drawing reliable conclusions from these data.

Change may not be relevant to patients unless it leads to "good" or "bad" outcomes. Good clinical outcome, defined as reaching a PASS, was positively associated with baseline hand function and negatively associated with the "identity" scale of the IPQ, meaning that patients who attribute fewer symptoms to hand OA are more likely to reach a PASS. These determinants are associated with both change in pain and with patient satisfaction after four years, emphasizing their importance. Being at PASS after four years was also associated with a lower tender joint count at baseline after adjustment for baseline pain, age,

sex, and BMI, indicating the number of affected joints may independently influence patient satisfaction measured through pain.

No other associations between change in pain or pain outcome and disease characteristics (Kellgren-Lawrence score, presence of erosions, synovitis, and symptom duration) were found after adjustment. This indicates that a cross-sectional association with pain, which is known to exist for erosive disease, does not necessarily indicate an association with change in pain, as well.[1,33] The discordance between radiographic damage and pain in hand OA has been described previously, and is again confirmed by our study.[2] Regarding symptom duration, no association was expected given the lack of change in pain over time on the group level. Change in synovitis measured on MRI was previously shown to be associated with change in pain on the joint level.[17] Previous ultrasonography studies have also shown that the association between synovitis and pain is stronger on the joint level than on the patient level.[6] In the current study, we found no effect between synovitis and pain on the patient level, indicating the effect may have been diluted because of the large number of joints contributing to the pain on patient level. This is exacerbated by the fact that synovitis was only scored in eight joints (the PIP and DIP joints of the right hand), which is a limitation in our study.

We used data from a large cohort, consisting of consecutively referred patients with hand OA presenting at the rheumatology outpatient clinic, only excluding patients with secondary OA or hand symptoms due to other causes. This sample is therefore expected to be representative of patients with hand OA seeking care from a rheumatologist.

There were a few limitations to our study in addition to the ones mentioned previously. Patients recruited from a secondary or tertiary center may be only partly generalizable to the larger hand OA population because not all patients with hand OA will visit a rheumatologist. Another limitation is that this study investigated the progression of hand OA pain in a cohort of patients with hand OA and could have been affected by collider stratification bias or selective loss to follow-up, most likely biasing found effects toward the null.[34] The real effects would then be larger than what was found in this study. As stated previously, this study could also suffer from residual confounding due to unmeasured variables, skewing the results in either direction. As such, replication and validation of these results is essential. We did not have data on all potential factors of interest, such as repeated hand movements. This is also a limitation of the current study that should be complemented in future studies.

To conclude, in this study we found that over four years, approximately 40% of patients will remain stable in their level of pain, with 30% of patients experiencing deterioration and 30% of patients experiencing improvement. The patients experiencing deteriorations are identified by higher BMI, having paid work, coping through comforting cognitions, and illness coherence. Patients experiencing improvements are identified by having paid work, mental well-being (less signs of anxiety and depression), more perceived consequences of hand OA, and better hand function. Over four years, the number of patients at a PASS slightly increased, which was associated with a lower number of tender joints and better hand function at baseline. These results can help inform patients and physicians. They may support the selection of patients for trials. These observational results require validation, but could represent modifiable risk factors, and require further study in future trials.

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 van der Meulen 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. Kloppenburg M, Kwok W-Y. Hand osteoarthritis--a heterogeneous disorder. Nat Rev Rheumatol 2011;8(1):22–31.
2. Marshall M, Watt FE, Vincent TL, et al. Hand osteoarthritis: clinical phenotypes, molecular mechanisms and disease management. Nat Rev Rheumatol 2018;14(11):641–656.
3. Steen Pettersen P, Neogi T, Magnusson K, et al. Peripheral and central sensitization of pain in individuals with hand osteoarthritis and associations with self-reported pain severity. Arthritis Rheumatol 2019;71(7):1070–1077.
4. O'Neill TW, Felson DT. Mechanisms of osteoarthritis (OA) pain. Curr Osteoporos Rep 2018;16(5):611–616.
5. Fu K, Robbins SR, McDougall JJ. Osteoarthritis: the genesis of pain. Rheumatology (Oxford) 2018;57(suppl 4):iv43–iv50.
6. Kortekaas MC, Kwok WY, Reijnierse M, et al. Pain in hand osteoarthritis is associated with inflammation: the value of ultrasound. Ann Rheum Dis 2010;69(7):1367–1369.
7. Gløersen M, Steen Pettersen P, Neogi T, et al. Associations of body mass index with pain and the mediating role of inflammatory biomarkers in people with hand osteoarthritis. Arthritis Rheumatol 2022;74(5):810–817.
8. Trouvin A-P, Perrot S. Pain in osteoarthritis. Implications for optimal management. Joint Bone Spine 2018;85(4):429–434.
9. Dahaghin S, Bierma-Zeinstra SM, Hazes JM, et al. Clinical burden of radiographic hand osteoarthritis: a systematic appraisal. Arthritis Rheum 2006;55(4):636–647.
10. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21(9):1145–1153.
11. Mulrooney E, Neogi T, Dagfinrud H, et al. The associations of psychological symptoms and cognitive patterns with pain and pain sensitization in people with hand osteoarthritis. Osteoarthr Cartil Open 2022;4(2):100267.
12. Bijsterbosch J, Watt I, Meulenbelt I, et al. Clinical and radiographic disease course of hand osteoarthritis and determinants of outcome after 6 years. Ann Rheum Dis 2011;70(1):68–73.
13. Vanhaverbeke T, Pardaens L, Wittoek R. Natural disease progression in finger osteoarthritis: results from a 10 year follow-up cohort. Scand J Rheumatol 2020;49(6):498–504.

14. Botha-Scheepers S, Riyazi N, Watt I, et al. Progression of hand osteoarthritis over 2 years: a clinical and radiological follow-up study. Ann Rheum Dis 2009;68(8):1260–1264.
15. van der Meulen C, van de Stadt LA, Rosendaal FR, et al. Determination and characterization of patient subgroups based on pain trajectories in hand osteoarthritis. Rheumatology (Oxford) 2023;62(9):3035–3042.
16. Kvien TK, Heiberg T, Hagen KB. Minimal Clinically Important Improvement/Difference (MCII/MCID) and Patient Acceptable Symptom State (PASS): what do these concepts mean? Ann Rheum Dis 2007;66 Suppl 3(Suppl 3):iii40–iii41.
17. van Beest S, Damman W, Liu R, et al. In finger osteoarthritis, change in synovitis is associated with change in pain on a joint-level; a longitudinal magnetic resonance imaging study. Osteoarthritis Cartilage 2019;27(7):1048–1056.
18. Damman W, Liu R, Kroon FPB, et al. Do comorbidities play a role in hand osteoarthritis disease burden? Data from the Hand OSTeoArthritis in Secondary Care cohort. J Rheumatol 2017;44(11):1659–1666.
19. Bellamy N, Campbell J, Haraoui B, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. Osteoarthritis Cartilage 2002;10(11):855–862.
20. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983;67(6):361–370.
21. Moss-Morris R, Weinman J, Petrie K, et al. The revised Illness Perception Questionnaire (IPQ-R). Psychol Health. 2002;17(1):1–16.
22. van Lankveld WNG, van der Staak C, van 't Pad Bosh P, et al. De ontwikkeling van de cors. Coping met reuma stressoren. Gedrag Gezond 1993;21:40–48.
23. 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.
24. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16(4):494–502.
25. Verbruggen G, Veys EM. Numerical scoring systems for the anatomic evolution of osteoarthritis of the finger joints. Arthritis Rheum 1996;39(2):308–320.
26. Haugen IK, Østergaard M, Eshed I, et al. Iterative development and reliability of the OMERACT hand osteoarthritis MRI scoring system. J Rheumatol 2014;41(2):386–391.
27. Altman R, Alarcón G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. Arthritis Rheum 1990;33(11):1601–1610.
28. Bellamy N, Hochberg M, Tubach F, et al. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. Arthritis Care Res (Hoboken) 2015;67(7):972–980.
29. Vanier A, Oort FJ, McClimans L, et al. Response shift in patient-reported outcomes: definition, theory, and a revised model. Qual Life Res 2021;30(12):3309–3322.
30. Liu R, Damman W, Kaptein AA, et al. Coping styles and disability in patients with hand osteoarthritis. Rheumatology (Oxford) 2016;55(3):411–418.
31. Damman W, Liu R, Kaptein AA, et al. Illness perceptions and their association with 2 year functional status and change in patients with hand osteoarthritis. Rheumatology (Oxford) 2018;57(12):2190–2219.
32. Kaptein AA, Bijsterbosch J, Scharloo M, et al. Using the common sense model of illness perceptions to examine osteoarthritis change: a 6-year longitudinal study. Health Psychol 2010;29(1):56–64.
33. Bijsterbosch J, Watt I, Meulenbelt I, et al. Clinical burden of erosive hand osteoarthritis and its relationship to nodes. Ann Rheum Dis 2010;69(10):1784–1788.
34. Zhang Y, Niu J, Felson DT, et al. Methodologic challenges in studying risk factors for progression of knee osteoarthritis. Arthritis Care Res (Hoboken) 2010;62(11):1527–1532.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 623–630
DOI 10.1002/acr.25464

American College of Rheumatology
Empowering Rheumatology Professionals

National Institute of Health and Care Excellence Clinical Criteria for the Diagnosis of Knee Osteoarthritis: A Prospective Diagnostic Accuracy Study in Individuals With Type 2 Diabetes

Lauren K. King,[1] Ian Stanaitis,[2] Vivian Hung,[2] Sahil Koppikar,[3] Esther J. Waugh,[4] Lorraine Lipscombe,[3] and Gillian A. Hawker[3]

Objective. The National Institute of Health and Care Excellence (NICE) criteria for osteoarthritis (OA) obviate the need for physical examination or imaging, and their use may improve timely diagnosis of OA. However, they have not been validated.

Methods. Within a larger study of individuals with type 2 diabetes, participants with and without self-reported knee pain underwent assessment of the NICE criteria for knee OA by questionnaire (index test) and clinical evaluation for established or possible knee OA by a rheumatologist (reference standard). We calculated the sensitivity, specificity, likelihood ratio positive (LR+), and likelihood ratio negative (LR–) of the NICE criteria and modified NICE criteria without the stiffness criterion.

Results. Our study included 96 participants: the mean ± SD age was 65.4 ± 8.3 years and 52% were women. Individuals who fulfilled the NICE criteria for knee OA (55.2%) included a spectrum of pain severity on an 11-point pain numeric rating scale with a median score of 5 (range 1–9). Rheumatologist assessment identified 56 participants (58.3%) with symptomatic knee OA. The sensitivity, specificity, LR+, and LR– of the NICE criteria for symptomatic knee OA were 0.84 (95% confidence interval [CI] 0.74–0.94), 0.85 (95% CI 0.74–0.96), 5.6, and 0.19, respectively. For the modified NICE criteria, these were 0.89 (95% CI 0.82–0.97), 0.85 (95% CI 0.74–0.96), 5.93, and 0.13.

Conclusion. The NICE criteria have high sensitivity and specificity for detecting symptomatic knee OA in a population with type 2 diabetes. We found that a modified version, omitting the stiffness criterion, performed similarly. These criteria should be validated in other settings and populations.

INTRODUCTION

Although knee osteoarthritis (OA) is a highly prevalent and disabling disease,[1] making a diagnosis of symptomatic knee OA, particularly across the spectrum of OA illness and disease severity, remains challenging for many clinicians. As a result, symptomatic knee OA is underdiagnosed and underdocumented by medical professionals,[2,3] which in turn limits patients' receipt of and engagement in OA care.[4] Making a diagnosis of knee OA requires the integration of patient history and examination.[5,6] Radiographs are not required for diagnosis[7,8] and in most cases, if acquired, do not change the diagnosis or management plan.[9–11] Radiographs alone cannot make a diagnosis of knee OA because a negative study could falsely rule out knee OA in an individual with early-stage preradiographic OA, and radiographic OA structural changes can be found in individuals without symptomatic

Supported by Physician Services Incorporated Foundation (resident research grant R21-28). The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or decision to submit the paper for publication.

[1]Lauren K. King, MD, PhD: University of Toronto and St. Michael's Hospital, Unity Health, Toronto, Canada; [2]Ian Stanaitis, BMSc, Vivian Hung, MPH: Women's College Hospital, Toronto, Canada; [3]Sahil Koppikar, MD, MSc, Lorraine Lipscombe, MD, MSc, Gillian A. Hawker, MS, MSc: University of Toronto and Women's College Hospital, Toronto, Canada; [4]Esther J. Waugh, BScPT, PhD: University of Toronto, Toronto, Canada.

All requests for access to anonymized data along with the data dictionary must be directed to Lauren K. King. Data transfer will be subject to an access protocol and will require authorization from Women's College Hospital Research Ethics Board. Data sharing will be contingent upon a research and data sharing agreement before any transfer, to ensure all users of the data adhere to the legal requirements of using personal data.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25464).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25464.

Address correspondence via email to Lauren K. King, MD, PhD, at l.king@mail.utoronto.ca.

Submitted for publication May 21, 2024; accepted in revised form November 5, 2024.

SIGNIFICANCE & INNOVATIONS

- The National Institute of Health and Care Excellence (NICE) guideline on osteoarthritis (OA) recommends that adults aged ≥45 years should be diagnosed with knee OA without investigation if they have activity-related knee pain and no morning joint-related stiffness or morning stiffness ≤30 minutes. Although the NICE criteria have been frequently used and referenced in the literature, there is limited literature with respect to their operating characteristics.
- In this diagnostic accuracy study, we found high sensitivity (84%) and specificity (85%) of the NICE criteria for knee OA. Performance was similar with the removal of the stiffness criterion. This suggests the modified version of the NICE criteria may be a preferable alternative.
- Given that they are simple to apply and obviate the need for physical examination or imaging, these criteria overcome existing barriers to making a diagnosis of knee OA for a wide range of clinicians.

knee OA.[12] The requirement for history and physical examination can be a barrier to making an OA diagnosis by clinicians less experienced in musculoskeletal assessments.[13] There is a need for a simple yet accurate OA diagnostic tool to identify individuals with knee OA within clinical care to improve the uptake of evidence-based treatment and enhance patient outcomes. Additionally, this need extends into research where it is valuable to have a pragmatic approach to the inclusion of individuals with knee OA into clinical studies.

The National Institute of Health and Care Excellence (NICE) guideline on OA recommends that adults aged ≥45 years should be diagnosed with OA clinically, without investigation, if they have activity-related knee joint pain and either no morning joint-related stiffness or morning stiffness that lasts ≤30 minutes.[8] These criteria were developed by expert consensus without validation studies. Three studies that have assessed operating characteristics of NICE criteria have been post hoc analyses of research data in selected populations,[14–16] with concern for both spectrum and reference standard bias. In a cross-sectional analysis of 13,459 participants of the Good Life with OA in Denmark (GLA:D) program with knee OA, the sensitivity of the NICE criteria compared with physical therapist assessment of knee OA was 89%.[14] In another study nested within the CHECK research cohort, Wang et al found that the NICE criteria had a sensitivity of 46% to 94% and specificity of 33% to 80%, depending on the number of follow-up time points (between one and three follow-up time points) in which participants fulfilled the criteria.[15] The study included individuals seeking first care for "knee complaints" in general practice and only those with complete follow-up data, and the reference standard for OA diagnosis was "expert consensus" based on a retrospective assessment of the research data. Finally, in a substudy of the ELSA-Brasil musculoskeletal study, the authors reported sensitivity of 57% and specificity of 73%. They included only participants with a knee radiograph within the past year, altered the morning stiffness criterion of the index test to ≤60 minutes, and they included the results of the radiograph into the reference standard assessment.[16] Thus, the accuracy of the NICE criteria in a general medical population remains unclear. Further, it is unknown whether the duration of morning joint stiffness ≤30 minutes is important to the functioning of the criteria given the known variability in joint stiffness reported by people living with knee OA.[17]

Our objective was to prospectively evaluate the accuracy of the NICE criteria for knee OA, and a modified version of the NICE criteria with the stiffness criterion removed, against a rheumatologist clinical assessment for knee OA. We hypothesized that the criteria would show high sensitivity and specificity for a rheumatologist diagnosis of knee OA and that removing the stiffness criterion would not change the accuracy. Understanding the operating characteristics of the NICE criteria are key to informing what role they play in the knee OA diagnostic pathway.

PATIENTS AND METHODS

Study design and participants. This was a diagnostic accuracy cross-sectional study nested within a larger multicenter, cross-sectional study of individuals with type 2 diabetes. For the larger study, we recruited participants age ≥45 years with type 2 diabetes who actively received care for type 2 diabetes (at least one visit within the past year) from endocrinology clinics at three academic hospitals in Toronto, Canada, between March 2022 and August 2023. We invited a subset of these participants with and without self-reported knee pain to participate (aiming for 50% with/without) in this diagnostic accuracy substudy. We invited consenting participants to attend a hospital outpatient appointment where they completed a self-administered questionnaire evaluating the NICE criteria (index test) immediately followed by clinical evaluation by a rheumatologist (reference standard). The near simultaneous assessment was to avoid biases caused by changes in participant disease status. A flow chart is shown in Figure 1. Demographic characteristics and medical history were assessed via online questionnaires as part of the larger cross-sectional study.

We received research ethics approval from the Women's College Hospital Research Ethics Board (REB #2020-0129-E) and the University of Toronto Office of Research Ethics (#40735). We followed the Standards for Reporting Diagnostic Accuracy reporting guideline.[18]

Index test: NICE criteria for knee OA. The NICE guideline on OA states that "adults aged ≥45 years should be diagnosed with OA clinically without investigations if they have

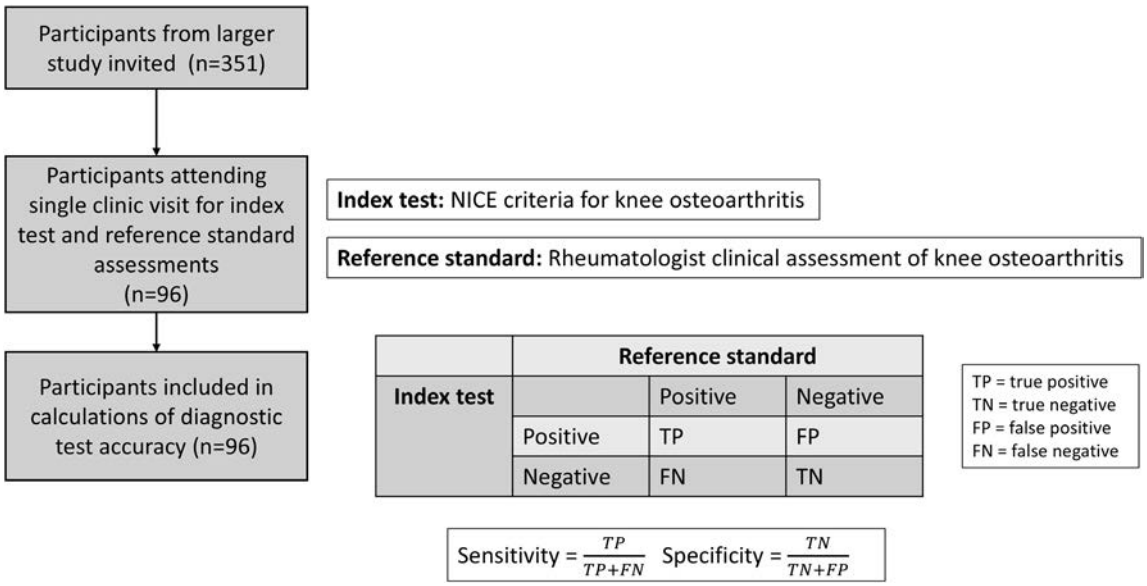


Figure 1. Flow chart of diagnostic accuracy study. NICE, National Institute of Health and Care Excellence.

activity-related joint pain and either no morning joint stiffness or morning stiffness that lasts less than 30 minutes."[8] The guidance advises against further evaluation with imaging to diagnose OA unless there are atypical features or features that suggest an alternative or additional diagnosis.[8]

To assess the NICE criteria for knee OA, participants self-completed a paper questionnaire that ascertained the presence of activity-related joint pain ("Do you have pain or aching in one or both of your knee joints that comes on, or is made worse by, activities such as standing, walking, or climbing stairs," yes/no) and the presence of morning joint stiffness ≤30 minutes ("Do you have stiffness in your knee(s) after waking up in the morning?", no/yes, up to 30 minutes/yes, 30 to 60 minute/yes, more than 60 minutes).

Reference standard: rheumatologist clinical diagnosis of knee OA. In a diagnostic test study, the test used as the benchmark (or gold standard) to evaluate the index test is called the reference standard.[19] We selected rheumatologist assessment of knee OA as the reference standard given that clinical history and physical examination is the preferred way to make a diagnosis of knee OA[5] and that rheumatologists are specialists in the diagnosis and management of arthritis. An experienced rheumatologist (LKK), blinded to participant responses to the index test but not to the purpose of the study, conducted a standardized clinical assessment to identify the presence of established knee OA (yes/no/possible). "Yes" indicated a clinical impression that the participant had established knee OA based on their clinical history (eg, knee joint symptoms, such as pain and aching and/or stiffness, and typical onset/course) and examination (eg, presence of joint-line tenderness, stress pain, and limitation in flexion and/or extension) without findings of an alternate diagnosis. "No" indicated a clinical impression that the participant did not have established knee OA. "Possible" indicated some residual uncertainty about the diagnosis after the single assessment in keeping with expected medical practice in which not all diagnoses can be determined based on a single visit. The reference standard was applied to all participants, irrespective of their response on the index test. As part of their assessment, participants were asked about the presence of diagnosed autoimmune inflammatory arthritis or other rheumatic or musculoskeletal disease. If a diagnosis other than knee OA was identified as part of the assessment, this was recorded by the rheumatologist. No diagnostic imaging was performed as part of the study.

A second rheumatologist (SK) completed a subset (n = 11) of assessments in duplicate to assess the reliability and reproducibility of the rheumatologist's assessment. We calculated the agreement (kappa) between the assessments of both rheumatologists.

Sample size. To detect an estimated sensitivity of 0.95, specificity of 0.85, prevalence of disease of 0.50 (among those assessed), precision of 0.10, and alpha of 0.05, we estimated that a sample size of 98 participants[20] would be required. Sensitivity and specificity estimates were selected based on the hypotheses of the research team.

Statistical analysis. We described the characteristics of the participants using mean ± SD, medians (interquartile range and/or range), and number (proportion), as appropriate. We calculated the proportion who fulfilled the NICE criteria for knee OA, defined as the presence of both activity-related knee joint pain and morning knee joint stiffness ≤30 min in individuals aged ≥45 years, and the modified NICE criteria, defined as the

presence of activity-related knee joint pain in individuals aged ≥45 years.

Sensitivity (the proportion of participants correctly identified by the index test as having the reference standard) and specificity (the proportion of participants correctly identified by the index test as not having the reference standard) are basic measures of the diagnostic accuracy of a test.[21] We calculated the sensitivity and specificity, with 95% confidence intervals (CIs), of the NICE criteria and modified NICE criteria to detect symptomatic knee OA (yes or possible). We considered yes or possible as positive given that these are individuals who should be picked up for further assessment and management. We also evaluated the likelihood ratio positive (LR+) and LR negative (LR–). These compare the probability that the index test will be positive (LR+) or negative (LR–) in someone with and without knee OA (yes or possible).

We generated receiver operator characteristic (ROC) curves to evaluate the performance of both the NICE and modified NICE criteria to discriminate individuals with rheumatologist assessed knee OA versus not. We also calculated the area under the ROC curve (AUC).

In our primary analysis, we apply the NICE criteria to all patients evaluated in an unselected manner. In a secondary analysis, we assessed the criteria only in those without a clinician diagnosis of another relevant chronic rheumatic and musculoskeletal disease (eg, rheumatoid arthritis).

To better understand the limitations of the NICE criteria, we used rheumatologist notes and results from the standardized clinical assessments to describe the characteristics of participants denoted as false positive or false negative and those identified as possible OA. We performed analyses using SAS Studio (version 3.81; SAS Institute Inc., Cary, NC).

RESULTS

Participants. Of 351 participants included in our larger study of people with type 2 diabetes, 96 participants consented to participate in the current substudy and underwent the index test and reference standard assessments. All 96 participants were included in our primary analyses (Figure 1). Their mean ± SD age was 65.4 ± 8.3 years, 51.6% were women, and the mean ± SD body mass index was 29.4 ± 6.6 kg/m^2 (Table 1). A total of 53 (55.2%) fulfilled the NICE criteria with a spectrum of illness severity: the median pain Numeric Rating Score (range 0–10) was 5 (range 1–9). A total of 56 (58.3%) fulfilled the modified NICE criteria. The reference standard assessment identified 56 (58.3%) participants with symptomatic knee OA (yes: n = 53 and possible: n = 3). We presented participant characteristics in Table 1.

Accuracy of the NICE criteria and modified NICE criteria. In our primary analysis of all 96 participants, the sensitivity and specificity of the NICE criteria for symptomatic knee OA were 0.84 (95% CI 0.74–0.94) and 0.85 (95% CI 0.74–0.96), respectively. The LR+ and LR– were 5.60 and 0.19, respectively. The sensitivity and specificity of the modified NICE criteria for symptomatic knee OA were 0.89 (95% CI 0.82–0.97) and 0.85 (95% CI 0.74–0.96), respectively. The LR+ and LR– were 5.93 and 0.13, respectively (Table 2). In our secondary analysis, after removing four individuals with physician diagnoses of chronic rheumatic and musculoskeletal disorders (rheumatoid arthritis [n = 1], psoriatic arthritis [n = 1], and fibromyalgia [n = 2]), the criteria performed similarly (Table 3).

Performance of NICE and modified NICE criteria. ROC curves for both the NICE criteria and modified NICE criteria are presented in Figure 2. The AUCs were as follows: NICE criteria 0.85 (95% CI 0.77–0.92) and modified NICE criteria 0.87 (95% CI 0.80–0.94). The curves were not statistically significantly different (P = 0.08).

Characteristics of false positives and negatives. False positives (ie, individuals who were positive for knee OA by the NICE criteria but negative for knee OA by rheumatologist assessment) included individuals with prior (self-limited) knee pain, recent knee trauma, fibromyalgia, and infrapatellar bursitis. Further details on the characteristics of the false positives are presented in Supplementary Table S2.

False negatives (ie, individuals who were negative for knee OA by the NICE criteria but positive for knee OA by rheumatologist assessment) included individuals who self-identified as having joint stiffness only and not "pain" (n = 2), those who reported >30 minutes of morning joint stiffness (n = 3), and an individual who described during clinical assessment that they tended to minimize their joint symptoms. Further details on the characteristics of the false negatives are presented in Supplementary Table 2.

Characteristics of "possible OA" and validation of rheumatologist assessment. We identified three individuals as having "possible OA" (ie, that additional information or longitudinal follow-up was required to make a diagnosis of established knee OA). For these individuals, history was compatible with knee OA; however, there were minimal clinical findings on knee examination. Clinician suspicion was for initial manifestations of knee OA. Further details on the characteristics of those with "possible OA" are presented in Supplementary Table 3. There was high rheumatologist interrater reliability for OA diagnosis (weighted κ = 0.86, 95% CI 0.60–1.00).

DISCUSSION

In this prospective diagnostic accuracy study, we found that the NICE criteria had high sensitivity and specificity for detecting symptomatic knee OA as assessed by an experienced clinician. Further, we found that a modified version without the stiffness

Table 1. Characteristics of participants (n = 96)*

Characteristic	Overall (n = 96)	Knee OA by rheumatologist	
		Yes (n = 56)	No (n = 40)
Age, mean (SD), years	65.4 (8.3)	65.4 (7.7)	65.5 (9.1)
Female sex, n (%)	50 (52.1)	34 (60.7)	16 (40.0)
Body mass index, mean (SD), kg/m^2	29.4 (6.6)	31.2 (7.1)	7.0 (5.0)
Race and ethnicity, n (%)			
Black	8 (8.3)	4 (7.1)	4 (10.0)
South Asian	7 (7.3)	6 (10.7)	1 (2.5)
East Asian	3 (3.1)	0 (0)	3 (7.5)
Southeast Asian	2 (2.1)	2 (3.6)	0 (0)
White	56 (58.3)	32 (57.1)	24 (60.0)
Middle Eastern	2 (2.1)	1 (1.8)	1 (2.5)
Other	16 (16.7)	9 (16.1)	7 (17.5)
Prefer not to answer	2 (2.1)	2 (3.6)	0 (0)
Diabetes treatment, n (%)			
Diet	51 (53.7)	29 (51.8)	23 (57.5)
Exercise	57 (59.4)	31 (55.4)	26 (65.0)
Oral medications	84 (87.5)	49 (87.5)	35 (87.5)
Insulin	50 (52.1)	29 (51.8)	21 (52.5)
Noninsulin injectable medications	31 (32.3)	21 (37.5)	10 (25.0)
Comorbidities, n (%)			
Stroke	9 (9.4)	5 (8.9)	4 (10.0)
Heart disease	18 (18.8)	11 (19.6)	7 (17.5)
Gastrointestinal disease	27 (28.1)	18 (32.1)	9 (22.5)
Kidney disease	17 (17.7)	12 (21.4)	5 (12.5)
Respiratory disease	22 (23.9)	15 (26.8)	7 (17.5)
Anxiety or depression	24 (25.0)	17 (30.4)	7 (17.5)
Established (physician diagnosis) chronic RMD, n			
Rheumatoid arthritis	1	1	0
Psoriatic arthritis	1	1	0
Fibromyalgia	2	1	1
Fulfilled the NICE criteria for knee OA, n (%)	53 (55.2)	50 (89.3)	6 (15.0)
Modified NICE criteria for knee OA, n (%)	56 (58.3)	47 (83.9)	6 (15.0)
Pain NRS (0–10) in those with knee pain, median (IQR)	5 (3–7)	5.5 (3–7) (n = 48 observations)	5 (5–5) (n = 5 observations)

* IQR, interquartile range; NICE, National Institute of Health and Care Excellence; NRS, Numeric Rating Score; OA, osteoarthritis; RMD, rheumatic and musculoskeletal disease.

criterion performed similarly and is a parsimonious alternative to the original NICE criteria. Given that the NICE criteria are easy to administer (two to three items, depending on whether the stiffness criterion is included), can be ascertained from self-report, and do not require physical examination or imaging, they are an important diagnostic tool for use within clinical practice and may be particularly valuable for clinicians, both physicians and non-physicians, who are not experts in musculoskeletal disease. The uptake of these criteria may increase the number of people with symptomatic knee OA who receive a clinician diagnosis, and, in turn, this may facilitate an increased uptake of evidence-based treatment.

This study provides important data on the accuracy of the NICE criteria in individuals with a spectrum of knee joint symptoms, from none to severe, who were not presenting for joint symptom evaluation and against the currently accepted gold standard for knee OA diagnosis. By virtue of this, the study overcomes limitations of prior studies that have evaluated the

Table 2. Accuracy of the NICE criteria (age ≥45 years, activity-related knee pain, and morning joint stiffness ≤30 min) and modified NICE criteria (age ≥45 years and activity-related knee pain) (n = 96)*

	Rheumatologist diagnosis		Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio	Negative likelihood ratio
NICE criteria	Positive	Negative	0.84 (0.74–0.94)	0.85 (0.74–0.96)	5.60	0.19
Positive	47	6				
Negative	9	34				
Modified NICE criteria	Positive	Negative	0.89 (0.82–0.97)	0.85 (0.74–0.96)	5.93	0.13
Positive	50	6				
Negative	6	34				

* CI, confidence interval; NICE, National Institute of Health and Care Excellence.

Table 3. Accuracy of the NICE criteria (age ≥45 years, activity-related knee pain, and morning joint stiffness ≤30 min) and modified NICE criteria (age ≥45 years and activity-related knee pain) in participants without a physician diagnosis of another RMD (n = 92)*

	Rheumatologist diagnosis		Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio	Negative likelihood ratio
NICE criteria	Positive	Negative	0.83 (0.73–0.93)	0.87 (0.77–0.98)	6.38	0.20
Positive	44	5				
Negative	9	34				
Modified NICE criteria	Positive	Negative	0.89 (0.80–0.97)	0.87 (0.77–0.98)	6.85	0.13
Positive	47	5				
Negative	6	34				

* CI, confidence interval; NICE, National Institute of Health and Care Excellence.

operating characteristics of the NICE criteria using retrospective data from research cohorts in individuals with joint symptoms.[14,15] The assessment of the modified NICE criteria further adds to the literature. The improved sensitivity of the modified NICE criteria suggested by our study may allow them to better function as a screening tool, whereby a nonmusculoskeletal expert could identify a patient as having possible OA and refer for further evaluation and management. The rationale for the requirement of ≤30 minute of morning joint stiffness to be diagnosed as OA remains unclear, as joint stiffness, even prolonged, has been identified by patients as a key component of the OA experience.[17] More general advantages of employing the NICE criteria include the ability to pick up individuals earlier in the disease course compared with the ACR classification criteria[22] and EULAR diagnostic criteria,[23] which select for individuals with more advanced symptoms, including functional impairment and joint structural changes.[24]

This study, through describing the false positives and negatives, indicates that the time frame of symptoms and terminology used in ascertaining the NICE criteria are likely important, and these are not specified in the NICE guidelines. In the current study, some false positives involved patients who had previously experienced knee symptoms (eg, due to fall) that had resolved. Specifying a time frame, such as symptoms experienced within last month, might improve specificity without compromising sensitivity. Further, we identified false negatives that may have been related to the terminology used. For the current study we asked participants, "Do you have pain or aching in one or both of your knee joints that comes on, or is made worse, by activities such as standing, walking, or climbing stairs?" This may not resonate with all people with OA, for example those who experience pain with kneeling while gardening. Future research should elucidate the optimal time frame and terminology.

In addition to the important role for the NICE criteria to ease identification of individuals with OA within clinical practice, the NICE criteria are increasingly being used as part of inclusion criteria in research studies. The high specificity of the NICE criteria and modified NICE criteria support their use for this purpose.

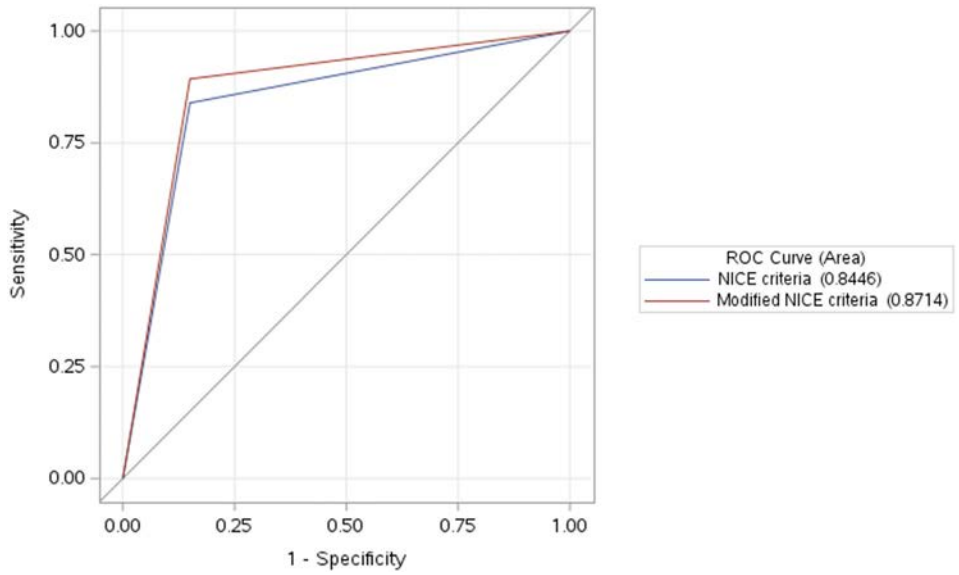


Figure 2. Paired ROC curves demonstrating the performance of the NICE criteria and modified NICE criteria to discriminate individuals with a rheumatologist diagnosis of knee osteoarthritis. NICE, National Institute of Health and Care Excellence; ROC, receiver operating characteristic.

These criteria present a pragmatic way to identify potentially eligible study participants compared with the ACR criteria,[22] which tend to identify individuals later in the illness and disease[24]; physician diagnosis, which limits to those who have sought and received medical care; or radiographs, which do not capture the illness of OA and reflect an established or late-stage disease process.

The study has many strengths. This is the first study, to our knowledge, to prospectively validate the NICE criteria for knee OA against a widely accepted reference standard. Participants included those with and without knee symptoms. Those with knee symptoms exhibited a broad spectrum of symptoms from mild to severe, capturing the entirely of the OA illness and disease, avoiding spectrum bias.[25] The participants undertook both the index and reference tests within a short time period. This reduced biases caused by changes in disease status, which can affect the diagnostic accuracy of the index test.[26] Included individuals had many other comorbidities, reflecting a medically complex population in which identifying OA may be especially important to prevent chronic disease complications.

This study has some limitations. We included only individuals with type 2 diabetes given that this was substudy. People with type 2 diabetes can develop vascular disease, neuropathy, and foot ulcers affecting the lower extremities, which could increase false positives. However, we did not identify these as reasons for false positives in our study. Nonetheless, the criteria should be validated in other settings and populations. Participants self-completed the questionnaire, and it is unknown how the operating characteristics might change if administered by a clinician. However, we suspect that, if anything, self-completion would bias toward lower sensitivity. Given purposeful recruiting (we aimed for 50% with and without knee pain), this study cannot assess positive and negative predictive values. Finally, some studies have suggested that there might be an inverse association between disease prevalence and specificity; given that the prevalence of knee OA in our study was higher than in the source population, the true specificity may be higher.[27]

In conclusion, in this diagnostic accuracy study completed in people with type 2 diabetes with and without knee pain, we found that the NICE criteria had high sensitivity (84%) and specificity (85%) for rheumatologist diagnosis of knee OA. The modified NICE criteria, in which the stiffness criterion was removed, performed similarly (sensitivity of 89% and specificity of 85%). Given its simplicity, we suggest that the modified NICE criteria may be a preferable diagnostic tool and may be particularly valuable to facilitate identifying knee OA in individuals within the care of other chronic diseases and by nonmusculoskeletal experts. Further research should confirm these results in different settings and populations and look to optimize and standardize terminology used and administration. The widespread use of an accurate way to identify people with knee OA within medical care, without the requirement for physical examination or imaging, has the potential to improve the number of individuals with symptomatic knee OA who receive a diagnosis and, by virtue of this, evidence-based treatment.

ACKNOWLEDGMENTS

The authors thank all patients for their participation in the study.

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Safiri S, Kolahi AA, Smith E, et al. Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020;79(6):819–828.
2. Arslan IG, Damen J, de Wilde M, et al. Incidence and prevalence of knee osteoarthritis using codified and narrative data from electronic health records: a population-based study. Arthritis Care Res (Hoboken) 2022;74(6):937–944.
3. Widdifield J, Jaakkimainen RL, Gatley JM, et al. Validation of Canadian health administrative data algorithms for estimating trends in the incidence and prevalence of osteoarthritis. Osteoarthr Cartil Open 2020; 2(4):100115.
4. Healey EL, Afolabi EK, Lewis M, et al. Uptake of the NICE osteoarthritis guidelines in primary care: a survey of older adults with joint pain. BMC Musculoskelet Disord 2018;19(1):295.
5. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet. 2019; 393(10182):1745–1759.
6. Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis: a review. JAMA 2021;325(6):568–578.
7. Sakellariou G, Conaghan PG, Zhang W, et al. EULAR recommendations for the use of imaging in the clinical management of peripheral joint osteoarthritis. Ann Rheum Dis 2017;76(9):1484–1494.
8. National Institute for Health and Clinical Excellence. Osteoarthritis in over 16s: diagnosis and management. NICE guideline NG226. Published October 19, 2022. Accessed February 24, 2024. https://www.nice.org.uk/guidance/ng226
9. Wang Q, Runhaar J, Kloppenburg M, et al; The CREDO Expert Group. The added value of radiographs in diagnosing knee osteoarthritis is similar for general practitioners and secondary care physicians; data from the CHECK Early Osteoarthritis Cohort. J Clin Med 2020;9(10):3374.
10. Mathiessen A, Cimmino MA, Hammer HB, et al. Imaging of osteoarthritis (OA): what is new? Best Pract Res Clin Rheumatol 2016;30(4): 653–669.
11. Skou ST, Thomsen H, Simonsen OH. The value of routine radiography in patients with knee osteoarthritis consulting primary health care: a study of agreement. Eur J Gen Pract 2014;20(1):10–16.
12. Jordan KP, Tan V, Edwards JJ, et al. Influences on the decision to use an osteoarthritis diagnosis in primary care: a cohort study with linked

survey and electronic health record data. Osteoarthritis Cartilage 2016;24(5):786–793.

13. Egerton T, Diamond LE, Buchbinder R, et al. A systematic review and evidence synthesis of qualitative studies to identify primary care clinicians' barriers and enablers to the management of osteoarthritis. Osteoarthritis Cartilage 2017;25(5):625–638.
14. Skou ST, Koes BW, Grønne DT, et al. Comparison of three sets of clinical classification criteria for knee osteoarthritis: a cross-sectional study of 13,459 patients treated in primary care. Osteoarthritis Cartilage 2020;28(2):167–172.
15. Wang Q, Runhaar J, Kloppenburg M, et al; CREDO expert group. Evaluation of the diagnostic performance of American College of Rheumatology, EULAR, and National Institute for Health and Clinical Excellence criteria against clinically relevant knee osteoarthritis: data from the CHECK Cohort. Arthritis Care Res (Hoboken) 2024;76(4):511–516.
16. Miguel RCC, Machado LA, Costa-Silva L, et al. Performance of distinct knee osteoarthritis classification criteria in the ELSA-Brasil musculoskeletal study. Clin Rheumatol 2019;38(3):793–802.
17. King LK, Mahmoudian A, Waugh EJ, et al; OARSI Early-stage Symptomatic Knee Osteoarthritis Initiative. "You don't put it down to arthritis": a qualitative study of the first symptoms recalled by individuals with knee osteoarthritis. Osteoarthr Cartil Open 2023;6(1):100428.
18. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015;351:h5527.
19. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. Acad Emerg Med 2013; 20(11):1194–1206.
20. Buderer NMF. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. Acad Emerg Med 1996;3(9):895–900.
21. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. BMJ 1994;308(6943):1552.
22. Altman R, Asch E, Bloch D, et al; Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis Rheum 1986;29(8): 1039–1049.
23. Zhang W, Doherty M, Peat G, et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010;69(3):483–489.
24. Peat G, Thomas E, Duncan R, et al. Clinical classification criteria for knee osteoarthritis: performance in the general population and primary care. Ann Rheum Dis 2006;65(10):1363–1367.
25. Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. Ann Intern Med 2002;137(7): 598–602.
26. Umemneku Chikere CM, Wilson K, Graziadio S, et al. Diagnostic test evaluation methodology: a systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - an update. PLoS One 2019;14(10):e0223832.
27. Murad MH, Lin L, Chu H, et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. CMAJ 2023;195(27):E925–E931.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 631–639
DOI 10.1002/acr.25481

American College of Rheumatology
Empowering Rheumatology Professionals

Risk of Incident Heart Failure and Heart Failure Subtypes in Patients With Rheumatoid Arthritis

Yumeko Kawano,[1] Brittany N. Weber,[1] Dana Weisenfeld,[2] Mary I. Jeffway,[2] Tianrun Cai,[2] Gregory C. McDermott,[1] Qing Liu,[2] Jeffrey A. Sparks,[1] Jennifer Stuart,[3] Jacob Joseph,[4] Tianxi Cai,[5] and Katherine P. Liao[6]

Objective. Patients with rheumatoid arthritis (RA) are at increased risk of cardiovascular disease (CVD) including heart failure (HF). However, little is known regarding the relative risks of HF subtypes such as HF with preserved ejection fraction (HFpEF) or reduced ejection fraction (HFrEF) in RA compared with non-RA.

Methods. We identified patients with RA and matched non-RA comparators among participants consenting to broad research from two large academic centers. We identified incident HF and categorized HF subtypes based on EF closest to the HF incident date. Covariates included age, sex, and established CVD risk factors. Cox proportional hazards models were used to estimate the hazard ratios (HRs) for incident HF and HF subtypes.

Results. We studied 1,445 patients with RA and 4,335 matched non-RA comparators (mean age 51.4 and 51.7 years, respectively; 78.7% female). HFpEF was the most common HF subtype in both groups (65% in RA vs 59% in non-RA). Patients with RA had an HR of 1.79 (95% confidence interval [CI] 1.38–2.32) for incident HF compared with those without RA after adjusting for CVD risk factors. Patients with RA had a higher rate of HFpEF (HR 1.99, 95% CI 1.43–2.77), but there was no statistical difference in the HFrEF rate (HR 1.45, 95% CI 0.81–2.60).

Conclusion. RA was associated with a higher rate of HF overall compared with non-RA, even after adjustment for established CVD risk factors. The elevated risk was driven by HFpEF, supporting a role for inflammation in HFpEF and highlighting potential opportunities to address this excess risk in RA.

INTRODUCTION

Patients with rheumatoid arthritis (RA) are at increased risk of cardiovascular diseases (CVD), such as myocardial infarction and stroke, compared with the general population.[1,2] Compared with ischemic heart disease, heart failure (HF) has been relatively understudied in RA despite significant contributions to morbidity and mortality.[3,4] Some prior studies have shown a heightened risk for HF in patients with RA compared with non-RA,[5,6] but HF was generally studied as a single entity and the differential risk of HF subtypes was limited by the need for large RA cohorts with detailed clinical documentation of phenotypic data (ie, echocardiograms, cardiology notes) required for such studies.

HF is a clinical syndrome characterized by symptoms, such as dyspnea, lower extremity edema, and evidence of congestion, that result from structural or functional abnormalities of the heart[7–9]; this definition encompasses both HF with reduced ejection fraction (HFrEF), defined as EF ≤40%, as well as HF with preserved ejection fraction (HFpEF) with EF ≥50%. Despite some overlap, there is increasing recognition of the different pathogenesis as well as treatment strategies of HFrEF compared with HFpEF.[10,11] Although HFrEF is typically caused by myocardial ischemic injury, HFpEF is thought to be associated with

Supported by NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grants R01-HL-127118 and P30-AR-072577). Dr Kawano's work was supported by the NIH Ruth L. Kirschstein Institutional National Research Service Award (grant T32-AR-007530).

[1]Yumeko Kawano, MD, Brittany N. Weber, MD, PhD, Gregory C. McDermott, MD, MPH, Jeffrey A. Sparks, MD, MMSc: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [2]Dana Weisenfeld, MS, Mary I. Jeffway, BS, Tianrun Cai, MD, Qing Liu, MS: Brigham and Women's Hospital, Boston, Massachusetts; [3]Jennifer Stuart, ScD: Brigham and Women's Hospital and Harvard T. H. Chan School of Public Health, Boston, Massachusetts; [4]Jacob Joseph, MBBS, MD: Veteran's Affairs Healthcare System, Boston, Massachusetts, and Veterans Affairs Healthcare System and Brown University, Providence, Rhode Island; [5]Tianxi Cai, ScD: Harvard T. H. Chan School of Public Health and Veteran's Affairs Healthcare System, Boston, Massachusetts; [6]Katherine P. Liao, MD, MPH: Brigham and Women's Hospital, Harvard Medical School and Veteran's Affairs Healthcare System, Boston, Massachusetts.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25481).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25481.

Address correspondence via email to Katherine P. Liao, MD, MPH, at kliao@bwh.harvard.edu.

Submitted for publication June 22, 2024; accepted in revised form December 2, 2024.

SIGNIFICANCE & INNOVATIONS

- This study uses a large cohort of patients with and without rheumatoid arthritis with detailed electronic health record data and leverages natural language processing to extract ejection fraction to examine the association between rheumatoid arthritis and the risk of specific heart failure subtypes.
- Patients with rheumatoid arthritis are at higher risk for heart failure and, in particular, heart failure with preserved ejection fraction compared with those without rheumatoid arthritis.
- Our results support a role for inflammation in the pathogenesis of heart failure with preserved ejection fraction and highlight opportunities to study the role of anti-inflammatory therapy to inform prevention and treatment of this condition in those with rheumatoid arthritis and in the general population.

pro-inflammatory metabolic conditions including obesity, diabetes, and potentially systemic inflammatory conditions such as RA.[12–14] Abnormal coronary microvascular dysfunction (CMD) caused by endothelial inflammation has been observed in HFpEF and may be one of the mechanisms by which HFpEF develops in RA.[15] However, despite a rising incidence in the general population, HFpEF remains relatively underdiagnosed in part because of its more subtle preclinical stages.[11] Understanding the risk of HF and specific HF subtypes in RA has the potential to improve screening, management, and ultimately cardiovascular outcomes for patients with RA.

The objective of our study was to investigate the risk of HF and HF subtypes, particularly HFpEF, between patients with RA compared with those without RA. As a prototypical inflammatory disease, we hypothesized that RA would be associated with increased risk of HF overall, as observed in prior studies, and specifically with HFpEF. We also examined the differences in risk factors for HFpEF in RA compared with non-RA as an exploratory analysis.

METHODS

Study design and data source. We conducted a retrospective cohort study using data from the Mass General Brigham (MGB) Biobank, a cohort of participants recruited from two large academic care centers who consented to broad-based research. The MGB Biobank is linked with electronic health record (EHR) data. The EHR data include both structured data, such as *International Classification of Diseases* (ICD) codes and laboratory values, as well as unstructured data, including clinical notes and imaging reports. The narrative data were extracted using natural language processing (NLP) with the Narrative Information Linear Extraction package.[16]

Patient population. Patients with RA were classified using a previously validated RA algorithm that combines ICD codes and RA-related NLP concepts (terms related to the phenotype of interest) with a positive predictive value (PPV) of 91%.[17] The index date for each patient with RA was the date of the first RA ICD code if followed by another RA ICD code within 6 months or the first RA NLP concept, whichever was earlier. Compared with a gold standard set of patients with RA for whom the diagnosis date was confirmed with manual medical record review, this rule for the RA diagnosis date performed best with an intraclass correlation coefficient of 0.88. We also required at least 90 days between the date of the first encounter in the EHR and RA diagnosis date (Figure 1) to identify patients with incident RA.

Each patient with RA was matched to three comparators without RA enrolled in the MGB Biobank. The comparators were matched based on sex, the year of birth (within the same year), and year of entry into the EHR (± 2-year window). The index date for each matched comparator was defined as an encounter date closest to the matched patient with RA's index date. Patients who were aged <18 years at the index date and those with <90 days of EHR history following the index date were excluded, as were patients with prevalent HF diagnoses before and up to 90 days following the index date (Figure 1).

Incident HF and identification of HF subtypes. Incident HF was ascertained using a previously validated algorithm (PPV 0.90) that incorporates a combination of structured and unstructured data from the EHR; the final algorithm included a weighted equation of the number of HF ICD codes, HF NLP, furosemide NLP, and total ICD counts.[18] The incident HF date was defined as the later date between the first HF ICD and the first furosemide NLP dates, as described in our prior study.[18]

We used the HF definition and classification criteria adopted by the American Heart Association, American College of Cardiology, Heart Failure Society of America, and the European Society of Cardiology.[9,10] HFpEF was defined as HF with an EF ≥50%, HFrEF was defined as HF with an EF ≤40%, and EF values in the 41%–49% range were classified as HF with moderately reduced EF (HFmrEF)[7] (Figure 1). HF subtypes were categorized using Extraction of Electronic Medical Record Numerical Data,[19] an NLP tool that extracts numerical data (ie, EF) from clinical notes and cardiology reports (including echocardiograms, cardiac magnetic resonance imaging scans, and cardiac positron tomography [PET] scans). The EF value closest to the HF incident date was used in the analysis. Those with HFmrEF were included in the overall HF outcome but not specifically studied as a separate outcome because of small numbers.

Follow-up started from 90 days after the index date to the earliest of incident HF (any type) to reduce the potential for

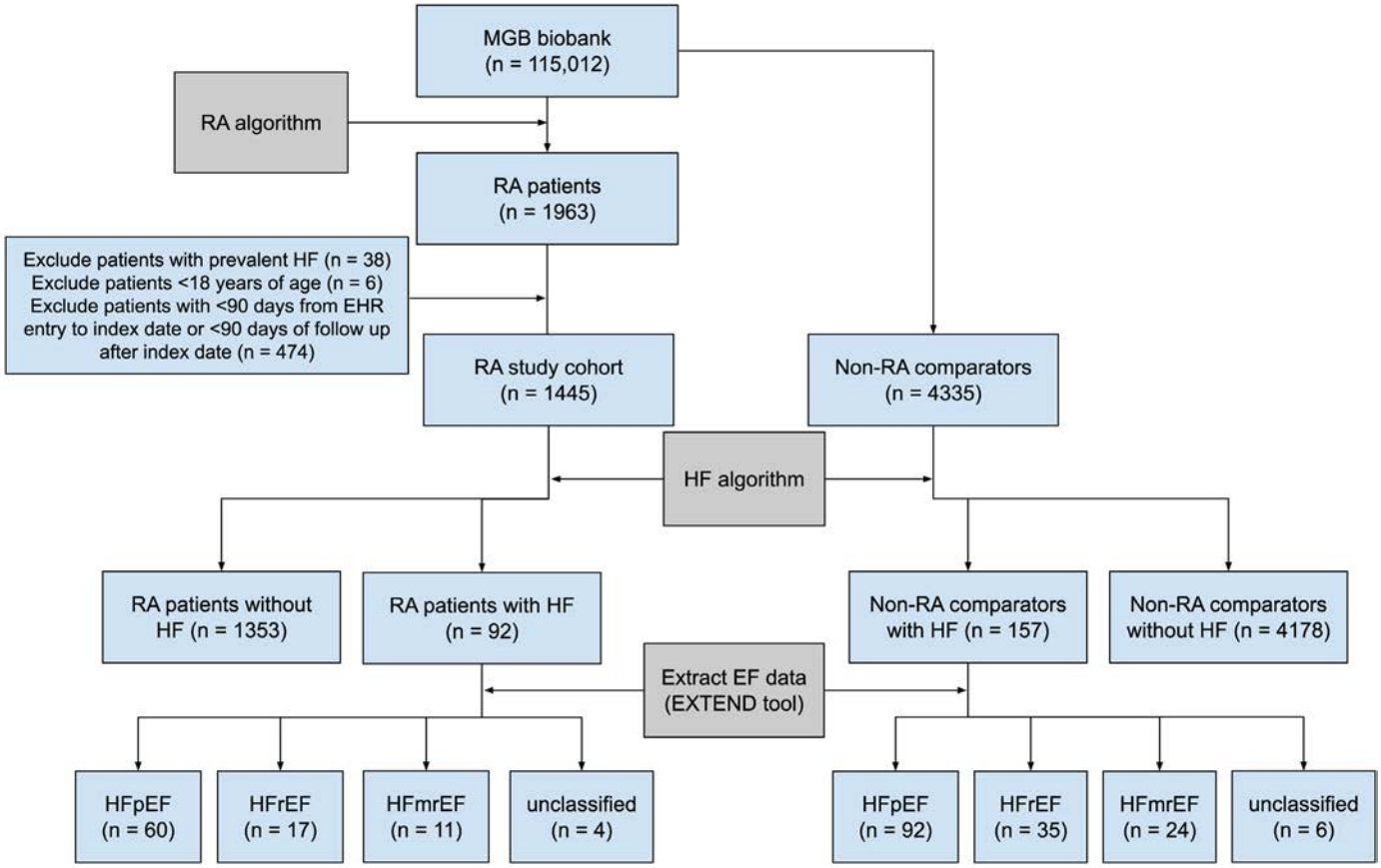


Figure 1. Study cohort selection of patients with RA and non-RA comparators. Patients with RA were first identified using a validated algorithm for RA. Each patient with RA was matched with three non-RA comparators matched by birth year, sex, and EHR entry date (±2 years). The HF algorithm was used to identify incident HF, and EXTEND was applied to extract EF from imaging studies and clinical notes closest to the HF incident date within 6 months. EF, ejection fraction; EHR, electronic health record; EXTEND, Extraction of Electronic Medical Record Numerical Data; HF, heart failure; HFmrEF, heart failure with moderately reduced ejection fraction; HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; MGB, Mass General Brigham; RA, rheumatoid arthritis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25481/abstract.

prevalent HF at the index date. Patients were observed until their last encounter in the EHR, death, end of study (October 26, 2021), or 15 years, whichever occurred earliest (Figure 2).

Covariates. We extracted demographic data and traditional cardiovascular risk factors. These included age at index date, sex, self-reported race, body mass index (BMI), smoking

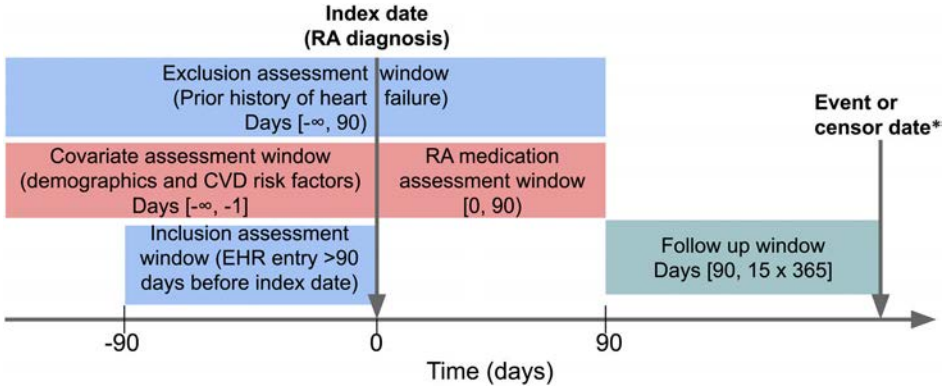


Figure 2. Overall study design to evaluate incident heart failure in patients with RA and non-RA comparators. *Censoring occurred at the earliest of incident heart failure (of any type), death, last EHR encounter, end of study (October 26, 2021), or 15 years. CVD, cardiovascular disease; EHR, electronic health record; RA, rheumatoid arthritis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25481/abstract.

status, history of coronary artery disease (CAD), atrial fibrillation, hypertension, hyperlipidemia, diabetes, stroke, and chronic kidney disease at index date for RA and non-RA comparators. Smoking status was classified as ever smoker versus never smoker based on positive mentions of smoking NLP concepts at or before the index date. Other comorbidities were ascertained based on ICD-9 and 10 codes as used in prior studies[6,18,20,21] (Supplementary Table 1).

For the RA cohort, we also extracted RA-specific variables including seropositivity (positive rheumatoid factor and/or antibodies to cyclic citrullinated peptide or NLP for seropositivity) at any point during the follow-up period. We also determined the use of corticosteroids and disease-modifying antirheumatic drugs (DMARDs) at baseline (within 90 days following the index date) (Figure 2). Postbaseline covariates, such as subsequent DMARD choice, nonsteroidal anti-inflammatory medication use, and disease activity, were not collected but may mediate the associations with outcomes.

Statistical analysis. The exposure of interest was RA versus non-RA status. Summary statistics for baseline characteristics of the cohort were reported as mean ± SD, median (interquartile range), or count (percentage), as appropriate; between-group comparisons were made using *t*-tests or chi-square tests.

Incident rates for HF overall and specific HF subtypes (HFpEF and HFrEF) were reported per 1,000 person-years. Time to incident HFpEF and HFrEF outcomes were assessed and displayed with cumulative incidence curves treating death as a competing event. We constructed unadjusted and adjusted multivariable Cox proportional hazards models to estimate the hazard ratios (HRs) for HF overall, HFpEF, and HFrEF. The models were adjusted by covariates selected a priori and included age, sex, and traditional CVD risk factors associated with HF: history of CAD, atrial fibrillation, hypertension, hyperlipidemia, diabetes, stroke, chronic kidney disease, and BMI. We used a cause-specific hazards model to examine the HF subtype outcomes (HFpEF and HFrEF), censoring patients who developed other HF subtypes at their HF diagnosis date. As an exploratory analysis, we stratified by RA status to enable a comparison of important clinical risk factors for HF and HFpEF between patients with and without RA.

We performed several sensitivity analyses. We examined a more stringent incident RA cohort using a longer run-in period of at least 1 year between the EHR entry date and the first RA code. We also used a Fine-Gray subdistribution hazard model as an alternate model to account for competing risk for the different HF subtypes. All analyses were performed using R version 4.2.1 (http://www.r-project.org/). This study was approved by the MGB institutional review board.

RESULTS

Among 1,963 patients who had incident RA, 1,445 met our inclusion and exclusion criteria (Figure 1), and they were matched in a 1:3 ratio to 4,335 comparators without RA. Baseline characteristics are shown in Table 1. The mean ± SD age was 51.4 ± 13.5 years for RA and 51.7 ± 13.4 years for non-RA cohorts, and 78.7% of both cohorts were female. Baseline comorbidities were similar between the two cohorts, except for CAD, which was more common in RA (10.9% in RA vs 8.5% in non-RA). The mean ± SD BMI was also slightly higher in RA (28.7 ± 6.7 kg/m^2 vs 28.1 ± 6.9 kg/m^2), and patients with RA were more likely to have ever smoked (30.9% vs 27.3%).

Over a combined 57,445 person-years of follow-up (median 10.3 years per patient), we identified 92 incident HF cases in RA and 157 incident HF in the non-RA cohorts (Table 2). HFpEF was the predominant HF subtype in both cohorts, accounting for 65.2% and 58.6% of all HF cases in RA and non-RA cohorts, respectively. The overall HF incidence rate was higher in RA than in non-RA comparators (6.63 vs 3.60 per 1,000 person-years). The incidence rate for HFpEF was higher in RA compared with

Table 1. Baseline characteristics of patients with RA and non-RA comparators in the Mass General Brigham Biobank*

Variable	Patients with RA (n = 1,445)	Non-RA comparators (n = 4,335)
Age, mean (SD), y	51.4 (13.5)	51.7 (13.4)
Female sex, n (%)	1,137 (78.7)	3,411 (78.7)
Race, n (%)		
Asian	33 (2.3)	107 (2.5)
Black or African American	97 (6.7)	223 (5.1)
Other	52 (3.6)	124 (2.9)
White	1,204 (83.3)	3,746 (86.4)
Unknown	59 (4.1)	135 (3.1)
Hispanic ethnicity, n (%)	38 (2.6)	91 (2.1)
Hypertension, n (%)	381 (26.4)	1,096 (25.3)
Hyperlipidemia, n (%)	396 (27.4)	1,186 (27.4)
Diabetes, n (%)	130 (9.0)	347 (8.0)
Chronic kidney disease, n (%)	52 (3.6)	127 (2.9)
Atrial fibrillation, n (%)	37 (2.6)	125 (2.9)
Coronary artery disease, n (%)[a]	157 (10.9)	367 (8.5)
Stroke, n (%)	80 (5.5)	228 (5.3)
Ever smoker, n (%)[a]	447 (30.9)	1,185 (27.3)
BMI, mean (SD),[a] kg/m^2	28.7 (6.7)	28.1 (6.9)
RA-related factors		
Seropositive	853 (59.0)	–
Baseline medications[b]		
Glucocorticoids	720 (49.8)	–
Methotrexate	573 (39.7)	–
Other csDMARDs	489 (33.8)	–
TNF inhibitors	159 (11.0)	–
Other b/tsDMARDs	39 (2.7)	–

* Patients with RA were matched to three non-RA comparators based on age, sex, and year of entry into the electronic health record (±2 years). All baseline characteristics were well matched in the two cohorts with some exceptions: a higher proportion of patients with RA had preexisting coronary artery disease, had ever smoked, and had higher baseline BMI. BMI, body mass index; b/tsDMARD, biologic and targeted synthetic disease-modifying antirheumatic drug; csDMARD, conventional synthetic DMARD; RA, rheumatoid arthritis; TNF, tumor necrosis factor.
[a] $P < 0.01$.
[b] Medications prescribed in the electronic health record within 90 days of the index date.

Table 2. Risk of incident HF and HF subtypes in patients with RA and non-RA comparators*

Summary data	Patients with RA (n = 1,445)	Non-RA comparators (n = 4,335)
Total follow-up time, person-years	13,877	43,568
Median follow-up time (IQR), years	9.6 (5.7–15.0)	10.5 (6.2–15.0)
Overall HF		
Cases, n	92	157
Incidence rate per 1,000 person-years	6.63	3.60
Unadjusted hazard ratio (95% CI)	1.84 (1.42–2.38)[a]	1.00 (reference)
Adjusted hazard ratio[b] (95% CI)	1.79 (1.38–2.32)[a]	1.00 (reference)
HF with preserved EF[c]		
Cases, n	60	92
Incidence rate per 1,000 person-years	4.33	2.11
Unadjusted hazard ratio (95% CI)	2.05 (1.48–2.84)[a]	1.00 (reference)
Adjusted hazard ratio[b] (95% CI)	1.99 (1.43–2.77)[a]	1.00 (reference)
HF with reduced EF[d]		
Cases, n	17	35
Incidence rate per 1,000 person-years	1.23	0.80
Unadjusted hazard ratio (95% CI)	1.54 (0.86–2.74)	1.00 (reference)
Adjusted hazard ratio[a] (95% CI)	1.45 (0.81–2.60)	1.00 (reference)

* BMI, body mass index; CI, confidence interval; EF, ejection fraction; HF, heart failure; IQR, interquartile range; RA, rheumatoid arthritis.
[a] $p < 0.05$.
[b] Adjusted for age, sex, and cardiovascular disease risk factors: baseline history of coronary artery disease, diabetes, hypertension, hyperlipidemia, atrial fibrillation, chronic kidney disease, stroke, and BMI.
[c] Defined as having an EF ≥50%.
[d] Defined as having an EF ≤40%. Those with a moderately reduced EF of 41% to 49% were included in the overall HF but excluded from the HF subtypes above.

non-RA (4.33 vs 2.11 per 1,000 person-years). The incidence rate for HFrEF did not differ significantly between the two groups (1.23 vs 0.80 per 1,000 person-years), although the total number of HFrEF events were small. A total of 31 patients with RA and 138 non-RA patients died during follow-up. Cumulative incidence curves for HFpEF and HFrEF between patients with RA and non-RA comparators, adjusting for the competing risk of death, are shown in Figure 3.

In the unadjusted Cox proportional hazards model, RA status was associated with increased hazard of HF overall (HR 1.84, 95% confidence interval [CI] 1.42–2.38) and HFpEF (HR 2.05, 95% CI 1.48–2.84) but not HFrEF (HR 1.54, 95% CI 0.86–2.74) (Table 2). In the multivariable model adjusting for age, sex, and traditional CVD risk factors (history of CAD, atrial fibrillation, hypertension, hyperlipidemia, diabetes, stroke, chronic kidney disease, and BMI), RA status was significantly associated with increased risk of HF overall (HR 1.79, 95% CI 1.38–2.32). Examining the HF subtypes, RA status was significantly associated with HFpEF (HR 1.99, 95% CI 1.43–2.77) but not HFrEF (HR 1.45, 95% CI 0.81–2.60).

In the exploratory analysis stratifying by RA status to compare clinical risk factors for HF and the HF subtypes between RA and non-RA comparators, we found that traditional CVD risk factors such as older age, BMI, and CAD history increase the hazard for HF overall in both groups (Table 3). Although female sex was associated with lower risk of HF overall in non-RA, it did not have as strong a protective effect among patients with RA. With regard to incident HFpEF, older age and higher BMI increased the risk for HFpEF in both groups; diabetes and stroke were also associated with increased risk of HFpEF among patients with RA but not among non-RA comparators.

In the sensitivity analysis with a more stringent criteria for incident RA requiring at least 1 year of EHR history before the RA diagnosis date/index date, we identified 1,329 patients with RA matched to 3,987 non-RA comparators. Similar to the primary analysis, we found that the incidence rate of HF overall was higher in RA compared with non-RA (6.63 vs 3.92 per 1,000 person-years), as was the incidence rate for HFpEF (4.35 vs 2.69 per 1,000 person-years). In multivariable Cox proportional hazards models (adjusting for the same covariates as in the primary analysis), we found similar results to the primary analysis, with a higher rate of HF overall (HR 1.66, 95% CI 1.27–2.17) and HFpEF (HR 1.59, 95% CI 1.14–2.21) but not HFrEF (HR 1.47, 95% CI 0.77–2.81) in RA compared with non-RA (Supplementary Table 2).

Using a Fine-Gray subdistribution hazard model of HFpEF (treating other HF subtypes as competing risk), the results were similar to the primary cause-specific model. Patients with RA had a higher hazard of HFpEF (adjusted HR 1.93, 95% CI 1.38–2.68) but not HFrEF (HR 1.37, 95% CI 0.76–2.49) (Supplementary Table 3).

DISCUSSION

In this study investigating the association of RA and HF with a focus on HF subtypes, patients with RA were at increased risk for HF overall and HFpEF compared with matched non-RA comparators after adjusting for known risk factors related to CVD. RA was not found to be an independent risk factor for HFrEF, although our findings regarding HFrEF are limited by small numbers. This study builds on prior literature demonstrating RA as a risk factor for HF and provides further detail that the signal is largely driven by risk for HFpEF.

Our findings of an estimated 79% increased risk for HF in RA compared with non-RA are in line with prior epidemiologic studies.[22] In one large population-based study conducted in the pre-biologic era, the risk for HF in RA was estimated to be as high as HR of 1.87 (95% CI 1.47–2.39).[5] In a more contemporary cohort, the HR was 1.21 (95% CI 1.03–1.42).[6] Furthermore, in the

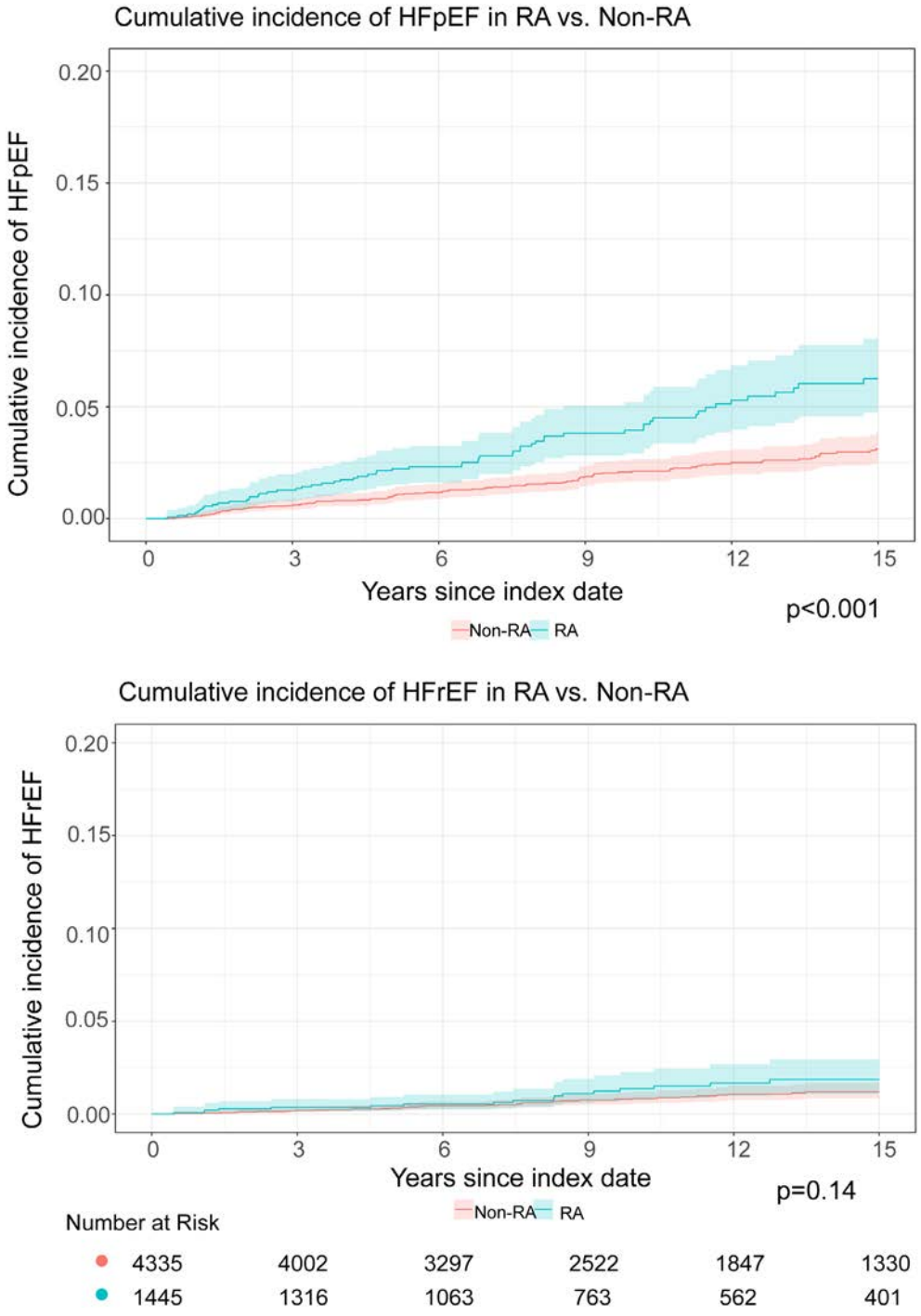


Figure 3. Cumulative incidence curves for HFpEF (top) and HFrEF (bottom) stratified by RA status, adjusting for the competing risk of death. HFpEF, heart failure with preserved ejection fraction; HFrEF, heart failure with reduced ejection fraction; RA, rheumatoid arthritis.

present study, the increased risk for HF in RA was driven by HFpEF rather than HFrEF. Few prior studies have examined HF subtypes specifically, likely because of the difficulty in identifying a sufficiently sized RA cohort with EF data in a format that can be used for large cohort analyses. Several studies have observed that patients with RA with HF tended to have higher EFs and evidence of diastolic dysfunction (one of the features of HFpEF) compared with non-RA patients with HF,[4,23] which aligns with our findings.

In one prior study, patients with RA were noted to have a heightened risk of nonischemic HF (defined by the absence of ICD codes for ischemic heart disease) compared with ischemic HF in the year following their RA diagnosis[24]; although EF values were not available in their study, ischemic heart disease is typically

Table 3. Factors associated with incident HF overall and incident HFpEF among RA and non-RA comparators*

Clinical factors	HF overall, HR (95% CI)		HFpEF, HR (95% CI)	
	RA	Non-RA	RA	Non-RA
Age	1.04 (1.02–1.07)[a]	1.06 (1.04–1.07)[a]	1.06 (1.03–1.09)[a]	1.07 (1.05–1.10)[a]
Female sex	0.70 (0.44–1.10)	0.69 (0.49–0.98)[a]	1.06 (0.58–1.94)	0.68 (0.43–1.06)
BMI	1.08 (1.05–1.11)[a]	1.04 (1.03–1.05)[a]	1.07 (1.04–1.11)[a]	1.05 (1.03–1.06)[a]
Hypertension	1.30 (0.74–2.29)	1.55 (1.02–2.36)[a]	1.57 (0.78–3.17)[a]	1.94 (1.13–3.35)[a]
Hyperlipidemia	0.72 (0.40–1.30)	0.76 (0.50–1.15)	0.72 (0.35–1.49)	0.72 (0.42–1.23)
Diabetes	1.72 (0.99–2.99)	1.64 (1.06–2.53)[a]	2.12 (1.09–4.12)[a]	1.07 (0.59–1.94)
Coronary artery disease	2.11 (1.18–3.78)[a]	2.32 (1.54–3.50)[a]	1.82 (0.89–3.73)	2.12 (1.26–3.57)[a]
Stroke	3.01 (1.65–5.48)[a]	1.57 (0.95–2.58)	2.75 (1.31–5.81)[a]	1.76 (0.97–3.21)
Atrial fibrillation	1.56 (0.71–3.42)	1.58 (0.92–2.72)	1.33 (0.49–3.64)	2.45 (1.36–4.43)[a]
Chronic kidney disease	1.19 (0.57–2.48)	1.06 (0.56–2.04)	1.19 (0.49–2.89)	1.22 (0.56–2.64)

* Multivariable Cox proportional hazards model for incident HF overall and HFpEF stratified by RA status. BMI, body mass index; CI, confidence interval; HF, heart failure; HFpEF, heart failure with preserved ejection fraction; HR, hazard ratio; RA, rheumatoid arthritis.
[a] $P < 0.05$.

associated with HFrEF. A recent Swedish registry-based study found that RA was more strongly associated with HF with EF ≥40% than EF <40% (odds ratio [OR] 1.7, 95% CI 1.4–2.0 vs OR 1.5, 95% CI 1.2–1.8)[25]; however, this association was driven more by HFmrEF (EF 40%–49%) than by HFpEF (EF ≥50%). Our study corroborates prior studies in which no increased risk for HFrEF was observed among the RA cases; however, the lack of signal in our study may also be due to a relatively small number of individuals with RA who developed an event.

HFpEF was the most common subtype of HF in our study, accounting for 65.2% of HF cases among patients with RA and 58.6% in the non-RA comparators. Prior studies have similarly noted that HFpEF was the predominant subtype among patients with RA and non-RA comparators[6,23]; this may be a reflection of the rising incidence of HFpEF in the general population,[11,26] but it nevertheless points to the importance of recognizing the HFpEF subtype in patients with RA. Some of the clinical risk factors for HFpEF in the general population include older age, higher BMI, atrial fibrillation, and diabetes mellitus.[27,28] Obesity and insulin resistance were particularly noted to be associated with HFpEF among women.[29] In the few studies investigating risk factors for HFpEF in the RA population, high levels of markers of inflammation (erythrocyte sedimentation rate and C-reactive protein) and high disease activity (Disease Activity Score-28) were noted to be associated with HFpEF.[6,18,23,24] HFpEF was the most common subtype of HF in our study, accounting for 66.3% of HF cases among patients with RA and 56.4% in the non-RA comparators. Prior studies have similarly noted that HFpEF was the predominant subtype among patients with RA as well as among non-RA comparators[6,23]; this may be a reflection of the rising incidence of HFpEF in the general population,[11,26] but it nevertheless points to the importance of recognizing the HFpEF subtype in patients with RA. Some of the clinical risk factors for HFpEF in the general population include older age, higher BMI, atrial fibrillation, and diabetes mellitus.[27,28] Obesity and insulin resistance were particularly noted to be associated with HFpEF among women.[29] In the few studies investigating risk factors for HFpEF in the RA population, high levels of markers of inflammation (erythrocyte sedimentation rate and C-reactive protein) and high disease activity (Disease Activity Score-28) were noted to be associated with HFpEF.[6,18,23,24] Our study builds on these studies and suggests that known cardiometabolic risk factors for HF, such as diabetes and obesity represented by higher BMI, conferred similar risk in RA and non-RA. BMI, however, has a complex relationship with RA and inflammation and may have a mediating role between RA and CVD. Furthermore, BMI has been found to have divergent effects on markers of inflammation based on sex in patients with RA,[30] and whether these sex differences translate into different CV risk profiles for men and women with RA requires further study.

Over the past decade, there has been increasing awareness of different phenotypes of HFpEF beyond diastolic dysfunction related to systemic hypertension. Particularly relevant to the rheumatology patient population is the inflammatory-metabolic phenotype, wherein systemic inflammatory states (such as obesity, diabetes, and systemic inflammatory diseases) are thought to induce CMD.[13,15,31] Specifically, proinflammatory cytokines including tumor necrosis factor α (TNFα), TNFα receptor 1, interleukin (IL) 6, IL-1, and others reduce endothelial production of nitric oxide, resulting in downstream effects that increase cardiomyocyte stiffening, hypertrophy, and myocardial fibrosis[32,33] and the clinical syndrome of HFpEF. Indeed, CMD as measured by impaired coronary flow reserve on cardiac PET scans have been observed in patients with RA without clinical CVD, especially in those with high IL-6 levels.[34,35] Whether CMD portends future HF risk (particularly HFpEF) and whether targeted anti-inflammatory therapies mitigate HFpEF risk in this population require further study. Nevertheless, minimizing comorbidities, such as diabetes and obesity, and controlling systemic inflammation in RA undoubtedly contribute to lowering the high CV risk burden in this population and should be pursued aggressively.[36]

Although HFpEF was once considered to have limited treatment options compared with HFrEF, the paradigm shifted after a landmark study in 2021 found that empagliflozin, a sodium-glucose cotransporter 2 inhibitor, reduced the risk of cardiovascular death and HF hospitalizations among patients with HFpEF.[37] Moreover, there has been increasing interest in the use of anti-inflammatory therapies for treatment of HFpEF. An analysis of the Canakinumab Anti-inflammatory Thrombosis Outcome Study (CANTOS) trial showed a dose-dependent decrease in HF hospitalizations with the use of the IL-1β monoclonal antibody canakinumab.[38] Following promising recent studies showing the association of IL-6 levels with incident HFpEF in the general (nonrheumatologic) population,[39,40] a large clinical trial is currently underway to study a novel IL-6 ligand monoclonal antibody (ziltivekimab) for the treatment of HFpEF and HFmrEF in the general cardiology population.[41] In the future, such studies could inform the treatment of patients with RA with HFpEF or cardiometabolic risk factors for HFpEF and highlight the importance of screening for HFpEF in this population.

Our study has several limitations. The study was conducted at an academic tertiary hospital system among participants in a Biobank, which may introduce selection bias as well as limitations in diversity and geography. Thus, the findings may not be generalizable to the general population. The study also did not account for postbaseline variables that may mediate associations, such as chronic nonsteroidal anti-inflammatory medication or steroid use, or specific DMARD use. Disease activity was also unavailable but likely mediates the association with HF outcomes. In addition, although the HF algorithm and the NLP method for extracting EF data were previously validated[18,19] and allowed us to effectively use EHR data, we relied on the availability of clinically performed cardiology studies for HF subtyping. Therefore, if patients received cardiology care or had cardiac imaging performed outside of the MGB system, we would not be able to ascertain their HF status or subtype. Additionally, although we sought to identify incident patients with RA by requiring 90 days of EHR history before any RA ICD code or NLP concept for RA, it is possible that there are patients whose RA diagnosis dates preceded the EHR entry date. We performed a sensitivity analysis using a more stringent criteria for incident RA, and the results were similar to those of the primary analysis. In our exploratory analysis comparing different risk factors for HF among RA and non-RA, we were underpowered to study HFrEF risk owing to the small number of events. Finally, CV comorbidities were ascertained using the common approach of using ICD codes; however, the PPVs can range from 71% to 95%.[6,18,20,21]

In summary, we observed that RA was an independent risk factor for incident HF, specifically HFpEF and not HFrEF, after adjusting for traditional CV risk factors associated with HF. Although the clinical risk factors for HFpEF did not vary substantially between RA and non-RA comparators, older age, higher BMI, and diabetes were associated with greater HFpEF risk in RA. RA can be considered a human model for inflammation, and findings from this study support the notion that chronic inflammation increases risk for HFpEF. Because inflammation is modifiable with anti-inflammatory medications, further studies are needed to determine whether anti-inflammatory therapies have the potential to reduce risk of HFpEF in those with RA and in other individuals with chronic inflammation.

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 Liao 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. Avina-Zubieta JA, Thomas J, Sadatsafavi M, et al. Risk of incident cardiovascular events in patients with rheumatoid arthritis: a meta-analysis of observational studies. Ann Rheum Dis 2012;71(9):1524–1529.
2. Solomon DH, Reed GW, Kremer JM, et al. Disease activity in rheumatoid arthritis and the risk of cardiovascular events. Arthritis Rheumatol 2015;67(6):1449–1455.
3. Nicola PJ, Crowson CS, Maradit-Kremers H, et al. Contribution of congestive heart failure and ischemic heart disease to excess mortality in rheumatoid arthritis. Arthritis Rheum 2006;54(1):60–67.
4. Davis JM III, Roger VL, Crowson CS, et al. The presentation and outcome of heart failure in patients with rheumatoid arthritis differs from that in the general population. Arthritis Rheum 2008;58(9):2603–2611.
5. Nicola PJ, Maradit-Kremers H, Roger VL, et al. The risk of congestive heart failure in rheumatoid arthritis: a population-based study over 46 years. Arthritis Rheum 2005;52(2):412–420.
6. Ahlers MJ, Lowery BD, Farber-Eger E, et al. Heart failure risk associated with rheumatoid arthritis-related chronic inflammation. J Am Heart Assoc 2020;9(10):e014661.
7. Bozkurt B, Coats AJS, Tsutsui H, et al. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure: endorsed by the Canadian Heart Failure Society, Heart Failure Association of India, Cardiac Society of Australia and New Zealand, and Chinese Heart Failure Association. Eur J Heart Fail 2021;23(3):352–380.
8. Yancy CW, Jessup M, Bozkurt B, et al; American College of Cardiology Foundation; American Heart Association Task Force on Practice Guidelines. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62(16):e147–e239.
9. Ponikowski P, Voors AA, Anker SD, et al; Authors/Task Force Members; Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special

contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2016;18(8):891–975.

10. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. Circulation 2022;145(18):e876–e894.
11. Borlaug BA, Sharma K, Shah SJ, et al. Heart failure with preserved ejection fraction: JACC scientific statement. J Am Coll Cardiol 2023; 81(18):1810–1834.
12. Redfield MM. Heart failure with preserved ejection fraction. N Engl J Med 2016;375(19):1868–1877.
13. Packer M, Lam CSP, Lund LH, et al. Characterization of the inflammatory-metabolic phenotype of heart failure with a preserved ejection fraction: a hypothesis to explain influence of sex on the evolution and potential treatment of the disease. Eur J Heart Fail 2020; 22(9):1551–1567.
14. Park E, Griffin J, Bathon JM. Myocardial dysfunction and heart failure in rheumatoid arthritis. Arthritis Rheumatol 2022;74(2):184–199.
15. Paulus WJ, Tschöpe C. A novel paradigm for heart failure with preserved ejection fraction: comorbidities drive myocardial dysfunction and remodeling through coronary microvascular endothelial inflammation. J Am Coll Cardiol 2013;62(4):263–271.
16. Yu S, Cai T, Cai T. NILE: fast natural language processing for electronic health records. arXiv Preprint posted online November 23, 2013. https://arxiv.org/abs/1311.6063
17. Huang S, Huang J, Cai T, et al. Impact of ICD10 and secular changes on electronic medical record rheumatoid arthritis algorithms. Rheumatology (Oxford) 2020;59(12):3759–3766.
18. Huang S, Cai T, Weber BN, et al. Association between inflammation, incident heart failure, and heart failure subtypes in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2023;75(5):1036–1045.
19. Cai T, Zhang L, Yang N, et al. EXTraction of EMR numerical data: an efficient and generalizable tool to EXTEND clinical research. BMC Med Inform Decis Mak 2019;19(1):226.
20. 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.
21. Kuang A, Xu C, Southern DA, et al. Validated administrative data based ICD-10 algorithms for chronic conditions: a systematic review. J Epidemiol Popul Health 2024;72(4):202744.
22. Khalid Y, Dasu N, Shah A, et al. Incidence of congestive heart failure in rheumatoid arthritis: a review of literature and meta-regression analysis. ESC Heart Fail 2020;7(6):3745–3753.
23. Schau T, Gottwald M, Arbach O, et al. Increased prevalence of diastolic heart failure in patients with rheumatoid arthritis correlates with active disease, but not with treatment type. J Rheumatol 2015; 42(11):2029–2037.
24. Mantel Ä, Holmqvist M, Andersson DC, et al. Association between rheumatoid arthritis and risk of ischemic and nonischemic heart failure. J Am Coll Cardiol 2017;69(10):1275–1285.
25. Faxén J, Benson L, Mantel Ä, et al. Associations between rheumatoid arthritis, incident heart failure, and left ventricular ejection fraction. Am Heart J 2023;259:42–51.
26. Tsao CW, Lyass A, Enserro D, et al. Temporal trends in the incidence of and mortality associated with heart failure with preserved and reduced ejection fraction. JACC Heart Fail 2018;6(8):678–685.
27. Lee MP, Glynn RJ, Schneeweiss S, et al. Risk factors for heart failure with preserved or reduced ejection fraction among Medicare beneficiaries: application of competing risks analysis and gradient boosted model. Clin Epidemiol 2020;12:607–616.
28. Ho JE, Lyass A, Lee DS, et al. Predictors of new-onset heart failure: differences in preserved versus reduced ejection fraction. Circ Heart Fail 2013;6(2):279–286.
29. Savji N, Meijers WC, Bartz TM, et al. The association of obesity and cardiometabolic traits with incident HFpEF and HFrEF. JACC Heart Fail 2018;6(8):701–709.
30. George MD, Giles JT, Katz PP, et al. Impact of obesity and adiposity on inflammatory markers in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2017;69(12):1789–1798.
31. Obokata M, Reddy YNV, Borlaug BA. Diastolic dysfunction and heart failure with preserved ejection fraction: understanding mechanisms by using noninvasive methods. JACC Cardiovasc Imaging 2020;13(1 Pt 2):245–257.
32. Paulus WJ, Zile MR. From systemic inflammation to myocardial fibrosis: the heart failure with preserved ejection fraction paradigm revisited. Circ Res 2021;128(10):1451–1467.
33. Shah SJ, Lam CSP, Svedlund S, et al. Prevalence and correlates of coronary microvascular dysfunction in heart failure with preserved ejection fraction: PROMIS-HFpEF. Eur Heart J 2018;39(37):3439–3450.
34. Amigues I, Tugcu A, Russo C, et al. Myocardial inflammation, measured using 18-fluorodeoxyglucose positron emission tomography with computed tomography, is associated with disease activity in rheumatoid arthritis. Arthritis Rheumatol 2019;71(4):496–506.
35. Weber B, Weisenfeld D, Seyok T, et al. Relationship between risk of atherosclerotic cardiovascular disease, inflammation, and coronary microvascular dysfunction in rheumatoid arthritis. J Am Heart Assoc 2022;11(11):e025467.
36. Peters MJ, Symmons DP, McCarey D, et al. EULAR evidence-based recommendations for cardiovascular risk management in patients with rheumatoid arthritis and other forms of inflammatory arthritis. Ann Rheum Dis 2010;69(2):325–331.
37. Anker SD, Butler J, Filippatos G, et al; EMPEROR-Preserved Trial Investigators. Empagliflozin in heart failure with a preserved ejection fraction. N Engl J Med 2021;385(16):1451–1461.
38. Everett BM, Cornel JH, Lainscak M, et al. Anti-inflammatory therapy with canakinumab for the prevention of hospitalization for heart failure. Circulation 2019;139(10):1289–1299.
39. Chia YC, Kieneker LM, van Hassel G, et al. Interleukin 6 and development of heart failure with preserved ejection fraction in the general population. J Am Heart Assoc 2021;10(11):e018549.
40. Markousis-Mavrogenis G, Tromp J, Ouwerkerk W, et al. The clinical significance of interleukin-6 in heart failure: results from the BIOSTAT-CHF study. Eur J Heart Fail 2019;21(8):965–973.
41. A research study to look at how ziltivekimab works compared to placebo in people with heart failure and inflammation (HERMES). ClinicalTrials.gov identifier: NCT05636176. Updated October 15, 2024. Accessed March 20, 2024. https://clinicaltrials.gov/ct2/show/NCT05636176

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 640–648
DOI 10.1002/acr.25467

American College of Rheumatology
Empowering Rheumatology Professionals

Immunosuppressive Drugs in Early Systemic Sclerosis and Prevention of Damage Accrual

Murray Baron,[1] Mandana Nikpour,[2] Dylan Hansen,[3] Susanna Proudman,[4] and Wendy Stevens,[5] on behalf of the Australian Scleroderma Interest Group and the Canadian Scleroderma Research Group, and Mianbo Wang[6]

Objective. Organ damage in patients with systemic sclerosis (SSc) in individual organs such as the lungs may be prevented by receiving immunosuppressive drugs (ISs). A new measure of global organ damage, the Scleroderma Clinical Trials Consortium Damage Index (SCTC-DI), has allowed us to investigate whether receiving ISs may reduce global organ damage accrual in patients with early SSc.

Methods. This was a retrospective study of patients with two or less years of disease duration in Canadian and Australian cohorts with SSc. Patients with either limited cutaneous SSc (lcSSc) or diffuse cutaneous SSc (dcSSc) were observed separately and divided into groups who were either ever or never exposed to ISs. The SCTC-DI was the outcome, and inverse probability of treatment weighting (IPTW) was used to balance the study groups and to fit a marginal structural generalized estimating equation model.

Results. In the cohort with lcSSc, there were 210 patients, of whom 34% were exposed to ISs at some time. Exposure to ISs was associated with lower damage scores. In the cohort with dcSSc, there were 192 patients, of whom 76% were exposed to ISs at some time. Exposure to ISs was not associated with damage scores.

Conclusion. In this retrospective observational cohort study, using IPTW to adjust for confounders, we found a protective effect of receiving ISs on damage accrual in patients with lcSSc. We were unable to determine such an effect in patients with dcSSc, but unknown confounders may have been present, and prospective studies of patients with dcSSc receiving ISs should include the SCTC-DI to determine the possible effect of ISs on damage accrual.

INTRODUCTION

Systemic sclerosis (SSc) is an autoimmune disorder characterized by vasculopathy, immunologic abnormalities, and fibrosis, affecting both skin and visceral organs. Although there is no cure, accumulating evidence shows that immunosuppression may be effective in stabilizing and perhaps improving manifestations of SSc, including skin thickening and interstitial lung disease (ILD).[1–11] Current recommendations support prescribing immunosuppressive drugs (ISs) for treatment of these manifestations of SSc.[9]

Organ damage is common and an important cause of mortality and morbidity in patients with SSc, and irreversible organ damage is accrued very early in the disease course, with 40% of patients having damage in one or more organ systems within two years of disease onset.[12–15] The relationship between immune abnormalities and organ damage is not completely clear, but studies have suggested that immunologic changes can ultimately lead

Supported by the Canadian Scleroderma Research Group (CSRG), which is funded by the Canadian Institutes of Health Research, Scleroderma Canada, the Scleroderma Society of Ontario, the Scleroderma Association of Saskatchewan, Scleroderma Manitoba, the Scleroderma Society of Nova Scotia, the Scleroderma Association of British Columbia, the Cure Scleroderma Foundation, the Canadian Blood and Marrow Transplant Group, and the Lady Davis Institute of Medical Research of the Jewish General Hospital, Montreal, Quebec, Canada. The CSRG has also received educational grants from Actelion pharmaceuticals and Mallinckrodt. The Australian Scleroderma Cohort Study is supported by research grants from Janssen, Boehringer Ingelheim, Australian Rheumatology Association, and Arthritis Australia. Dr Nikpour's work was supported by the National Health and Medical Research Council of Australia Emerging Leadership (investigator grant GNT-1176538).

[1]Murray Baron, BSc, MDCM, FRCP (C): Jewish General Hospital, McGill University, Montreal, Quebec, Canada; [2]Mandana Nikpour, MBBS, FRACP, FRCPA, PhD: The University of Sydney and Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia; [3]Dylan Hansen, MbioStat, Bbiomed, Sc(PharmSc): The University of Sydney, Camperdown, New South Wales, Australia; [4]Susanna Proudman, MBBS (Hons), FRACP: Royal Adelaide Hospital, Adelaide, South Australia, Australia; [5]Wendy Stevens, MBBS, FRACP: St. Vincent's Hospital Melbourne, Fitzroy, Victoria, Australia; [6]Mianbo Wang, MSc: Lady Davis Institute for Medical Research, Montreal, Quebec, Canada.

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.25467).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25467.

Address correspondence via email to Murray Baron, BSc, MDCM, FRCP (C), at murray.baron@mcgill.ca.

Submitted for publication April 26, 2024; accepted in revised form October 30, 2024.

SIGNIFICANCE & INNOVATIONS

- The effect of receiving immunosuppressive drugs in patients with systemic sclerosis (SSc) on global damage has not been assessed before.
- Using the Scleroderma Clinical Trials Consortium Damage Index and data from Canada and Australia, we assessed the effect of receiving immunosuppressive drugs in patients with SSc on global damage.
- In patients with early disease, we have found that receiving immunosuppressive drugs reduces damage accrual in patients with limited cutaneous SSc.
- More research is needed to assess the effect of receiving immunosuppressive drugs on global damage in patients with early diffuse cutaneous SSc.

to tissue fibrosis, a hallmark of disease damage in SSc.[16–20] It was previously difficult to quantitate organ damage accrual in SSc, but the development and validation of the Scleroderma Clinical Trials Consortium Damage Index (SCTC-DI) has opened the door to studies of damage accrual in this disease.[12–14]

The objective of this study was to assess the relationship between receiving ISs and the development of organ damage over time in a longitudinal observational cohort of patients with SSc. The ISs assessed were methotrexate (MTX), cyclophosphamide (CYC), azathioprine (AZA), and mycophenolate (MPA). Our hypothesis was that receiving ISs would prevent the development of damage compared to not receiving such medications. To adjust for imbalance in measured confounders between groups who were and were not exposed, inverse probability of treatment weighting (IPTW) was used to balance the characteristics of the two patient groups.

PATIENTS AND METHODS

Study populations. Participants were recruited from the Canadian Scleroderma Research Group (CSRG) registry and from the Australian Scleroderma Cohort Study (ASCS). The CSRG recruits and follows patients from 15 centers in Canada and Mexico. These centers see local and regional referrals. All patients must have a diagnosis of SSc (confirmed by an experienced rheumatologist), be ≥18 years of age, provide informed consent, and be fluent in English, French, or Spanish. Over 98% of the cohort meets the 2013 American College of Rheumatology (ACR)/EULAR classification criteria for SSc.[21] The ASCS is an Australian multicenter cohort study of risk and prognostic factors in SSc. The ASCS and CSRG have been approved by all human research ethics committees of participating sites. For a complete list of members of the Australian Scleroderma Interest Group and the CSRG, who collected data for this paper and reviewed the manuscript, see Appendices A and B. All participants met ACR/EULAR criteria for SSc.[21] Written informed consent was obtained from all participants at recruitment. We included only those participants recruited within two years of onset of the first non-Raynaud manifestation of SSc. Patients with diffuse cutaneous SSc (dcSSc) were defined by skin thickening proximal to the elbows or knees and/or trunk at any time. Ethics committee approval for this study was obtained at the Jewish General Hospital, Montreal, Canada (approval number 2019-1597) and at all participating CSRG study sites. Ethics approvals for Australian data were obtained from St. Vincent's Hospital Melbourne (approval number LRR 012/21) and all participating ASCS sites.

Exposure. Medication exposure was recorded yearly by study physicians and coded as current, past, or never. Patients were recorded at each visit as being exposed or not to ISs (MTX, CYC, AZA, or MPA). At the baseline visit either no IS, current IS, or IS not current but before the visit exposure was recorded. In our analysis, "immunosuppressants before baseline" refers to ISs not received at the baseline visit but only before that visit.

SCTC-DI. The SCTC-DI measures global irreversible damage in patients with SSc and was developed to be highly correlated with mortality and morbidity (measured by the 36-item short-form (SF-36) survey).[22] Validated and published in 2019, its development was an international collaboration among 22 experts with input from patient partners using a combined approach of consensus and data-driven methods. The index is composed of 23 differently weighted items in several organ systems (musculoskeletal, skin, vascular, gastrointestinal, respiratory, cardiovascular, and renal). Low, medium, and high SCTC-DI scores are defined as <5, 6 to 12, and ≥13, respectively, with a maximum score of 55. For the definition of the individual components of the SCTC-DI, please refer to the original paper.[22]

In this study, the SCTC-DI was calculated using registry data, but three items (calcinosis complicated by infection or requiring surgery, gastric antral vascular ectasia, and right ventricular dysfunction) were removed because they were not collected in the CSRG database, and one item (small-joint contractures) was removed due to missing data (>20% of visits). The maximum SCTC-DI score possible was therefore 42. Scores are always whole integers, as are changes in scores over time.

To score the SCTC-DI for lung disease, high-resolution computed tomography (HRCT) scans were assessed by local radiologists, and there was no central reading. CT scans were not necessarily done according to any protocol but were done at the request of the treating physician. Once a damage score was assigned, it is carried forward until something occurs to increase it. In the case of the lung, an increase in score would occur if another CT scan showed an increase of ILD on CT to ≥20% or a drop in forced vital capacity (FVC) to <70% predicted. The damage score is calculated from the data acquired at each patient visit. The SCTC-DI is scored such that patients cannot improve; they can

only stay the same or worsen, so unless there is worsening in a particular organ, the score remains the same as the previous visit.

Definition of variables. Disease duration was defined from the onset of the first non-Raynaud phenomenon symptom to the index visit (Table 1). Smoking status was classified as either never smoker or past and/or current smoker. Skin involvement was assessed using the modified Rodnan skin score (mRSS), which ranges from 0 (no involvement) to 3 (severe thickening) in 17 areas (score range 0–51). FVC percentage was extracted from pulmonary function tests. The presence of ILD was determined using a published clinical decision rule.[23] Using this rule, ILD was considered present if an HRCT scan of the lung was interpreted by an experienced radiologist as showing ILD or, in the case of no HRCT being available, if either a chest x-ray was reported as showing either increased interstitial markings (not thought to be due to congestive heart failure) or fibrosis and/or if a study physician reported the presence of typical "Velcro-like crackles" on physical examination. Physical function was assessed using the Health Assessment Questionnaire Disability Index, which is scored from 0 (no disability) to 3 (severe disability). Patient and physician global assessment scores were rated 0 to 10 (no disease to very severe disease) on numeric rating scales. For patient assessment scores, patients were asked, "In the past week, how was your overall health?" The physician global severity question asked, "How would you rate the patient's overall health for the past week?" Other covariates recorded at the index visit included physician reports of inflammatory myositis, arthritis, digital ulcers, and prior scleroderma renal crisis. Pulmonary hypertension was defined as an estimated systolic pulmonary artery pressure ≥45 mm Hg measured using the Doppler flow measurement of the tricuspid regurgitant jet on cardiac echocardiography (used as a noninvasive screening tool for pulmonary hypertension).[24] Antinuclear antibody was detected by immunofluorescence, and other autoantibodies were detected by line immunoassay (Euroimmun).

Statistical analysis. Because of the very different trajectories of damage in limited cutaneous SSc (lcSSc) and dcSSc,[12,13] and because exposure to ISs was so different in the subsets of patients with cutaneous SSc (145 of 192 patients or 76% in the subset with dcSSc and 71 of 210 or 34% in the subset with lcSSc), we analyzed these two subsets separately. Descriptive statistics were used to summarize baseline demographic and clinical characteristics of the patients who were and were not exposed to ISs. Continuous variables are presented as mean ± SD, and categorical variables are presented as counts and percentages. Student's *t*-test and Wilcoxon–Mann–Whitney U test were used to compare continuous variables. Chi-square test and Fisher exact test were used for categorical variables.

Due to the inherent differences between patients who were and were not exposed to ISs in an observational study, IPTW was used to balance the study groups and to fit a marginal structural

Table 1. Baseline characteristics of limited cutaneous scleroderma patients (lcSSc) stratified according to exposure to immunosuppression at baseline or during follow-up (*N* = 210)*

	Ever exposed to IS at any visit (*n* = 71)		Never exposed to IS at any visits (*n* = 139)		*P*-values
		Missing data *N*		Missing data *N*	
Age, years (mean ± SD)	55.1 ± 11.8		56.4 ± 13.7		0.509
Female, N (%)	60 (84.5%)		121 (87.1%)		0.613
Caucasian, N (%)	63 (92.6%)	3	113 (83.7%)	4	0.077
Education (>high school), *N* (%)	31 (45.6%)	3	60 (46.5%)	10	0.902
Smoking in the past or currently, *N* (%)	39 (54.9%)		71 (51.1%)		0.597
Disease duration, years (mean ± SD)	1.1 ± 0.5		1.0 ± 0.5		0.517
mRSS (0–51) (mean ± SD)	5.9 ± 4.3	1	4.4 ± 4.2	1	0.003
Interstitial lung disease, *N* (%)	23 (32.4%)		27 (19.4%)		0.037
FVC, % predicted (mean ± SD)	93.8 ± 20.8	0	97.5 ± 20.2	1	0.214
DLCO, % predicted (mean ± SD)	69.0 ± 20.6	4	73.3 ± 22.6	9	0.192
Tendon friction rubs, *N* (%)	4 (5.6%)	0	3 (2.2%)	2	0.233
Inflammatory arthritis, N (%)	34 (48.6%)	1	30 (22.2%)	4	<0.001
Autoantibodies					
Anti-centromere, *N* (%)	20 (30.0%)	2	79 (59.4%)	6	<.001
Anti-topoisomerase, *N* (%)	23 (33.3%)	2	13 (9.7%)	5	<.001
Anti-RNA polymerase III, *N* (%)	6 (9.1%)	5	9 (7.3%)	16	0.667
C-reactive protein, mg/L (median, IQR)	4 (2–7)	6	3.7(1.4–6.4)	9	0.109
Damage score at baseline (median, IQR)	3 (2–5)		3 (1–5)		0.139
CSRG patients, *N* (%)	32 (45.1%)		56 (40.3%)		0.506
Any Immunosuppressants prior to but not at baseline, *N* (%)	6 (8.5%)		4 (2.9%)		0.147

* CSRG, Canadian Scleroderma Research Group; DLCO, diffusing capacity of carbon monoxide; FVC, forced vital capacity; IS, immunosuppressives; mRSS, modified Rodnan skin score.

generalized estimating equation (GEE) model.[25] Each in-person visit was treated as an observation. Propensity scores representing the probability of being exposed at a given visit were calculated using a pooled logistic regression adjusting for baseline covariates (sex, age, disease duration, exposure to immunosuppression before baseline visit, FVC at baseline, damage score at baseline visit) and time-varying covariates (mRSS and presence of arthritis) at each annual visit. These time-varying variables not only affect treatment but also associate with the outcome, and they also may be affected by the previous treatment. They are potentially time-varying confounding, which is important to adjust for as time-varying variables in the propensity score model for the IPTW.

To evaluate for residual differences in baseline covariates between the two statistically matched groups, we calculated the standardized mean difference (SMD) of each disease variable. The SMD is the difference in means of a covariate across the treatment groups divided by the SD in the treated group. It is the most commonly used statistic to examine the balance of covariate distribution between treatment groups.[26,27] Because the SMD is independent of the unit of measurement, it allows comparison between variables with different units of measurement. A standardized difference of ≤0.1 represents meaningful balance.[26,27]

All models include time-varying variables. We have adjusted and unadjusted models for weights. The adjusted model was used to account for the time-varying variables and generated the denominator of the weights. The unadjusted model was used to generate the numerator of the weights, which created more stabilized weights. These stabilized weights were then used in a weighted GEE model to estimate the parameters of the marginal structural model. This model was conditional on exposure to ISs at the given visit and adjusted for sex, age, disease duration, diffuse subset, exposure to immunosuppression before baseline visit, and damage score at baseline visit. The cohort (CSRG vs ASCS) was also adjusted in the GEE outcome model. Because the damage scores are always increasing, we used the change of damage scores between two visits (eg, damage score at current visit minus damage score at previous visit) as the outcome at each visit instead of the raw scores. Because the most common antibodies in patients with lcSSc were anti-centromere and anti-topoisomerase and in patients with dcSSc were anti–RNA polymerase III and anti-topoisomerase, in the respective models we assessed those antibodies versus all other patients.

RESULTS

Supplementary Table 1 shows the ISs received at or before the baseline visit. Two patients received both MTX and MPA, so a total of 170 of 402 patients (42%) received some IS already at or before the baseline visit in a cohort of patients with SSc with a disease duration two or fewer years. Figure 1 shows the mean SCTC-DI values in each group over time. These figures represent unweighted changes, and it can be seen that patients with dcSSc who were immunosuppressed accrued damage at a faster rate than patients who were not immunosuppressed (Figure 1C); in patients with lcSSc, the rate of damage accrual is similar in both patients who were and were not immunosuppressed until about six years, after which patients who were immunosuppressed accrued damage faster (Figure 1B); however, the number of patients is small after that time point.

Cohort with lcSSc. There were 210 patients in the subset with lcSSc, of which 34% were exposed to ISs at some time (Table 1). Mean (±SD) follow-up was 4.3 (±2.9) years. Patients who had exposure at some point had higher mean mRSS (5.9 ± 4.3 vs 4.4 ± 4.2; $P = 0.003$), were more likely to have ILD (32.4% vs 19.4%; $P = 0.037$), were more likely to have inflammatory arthritis, were less likely to have anti-centromere antibody (30.0% vs 59.4%; $P < 0.001$), and were more likely to have anti-topoisomerase antibodies (33.3% vs 9.7%; $P < 0.001$). Very few patients received ISs only before baseline. The balance diagnostics of the covariates at baseline visit for patients with lcSSc after propensity scoring are illustrated in Supplementary Table 2. The balancing of all variables was better after IPTW, but the SMDs were suboptimal for disease duration and IS before baseline.

Table 2 shows the association between exposure to immunosuppression and the change of damage scores adjusted for possible confounders. Exposure to ISs was associated with lower damage scores. The coefficient for the effect of immunosuppression on the damage index is –0.34 after IPTW (Table 2). A coefficient of –0.34 means that the change in damage score from the previous visit will be 0.34 points less than the change when the patient did not receive ISs. If we take the values from Figure 1B as an estimate of change over time in patients with lcSSc who did not receive ISs, the mean change in score between yearly visits is about 0.3. So, to reduce those scores by 0.34 suggests that the effect of receiving ISs might be substantial. Lower FVC values were associated with more damage accrual, both with and without IPTW.

Cohort with dcSSc. There were 192 patients in the subset with dcSSc, of whom 76% were exposed to ISs at some time (Table 3). Mean (±SD) follow-up was 4.6 (±3.0) years. Patients who were exposed at some point were younger (50.0 ± 12.1 years versus 55.5 ± 10.7 years; $P = 0.008$), were less likely to have anti-centromere antibodies (3.0% vs 18%; $P = 0.001$), and were more likely to have anti-topoisomerase antibodies (33.1% vs 10.3%; $P = 0.005$). Very few patients received ISs only before baseline.

The balance diagnostics of the covariates at baseline visit for patients with dcSSc after propensity scoring are illustrated

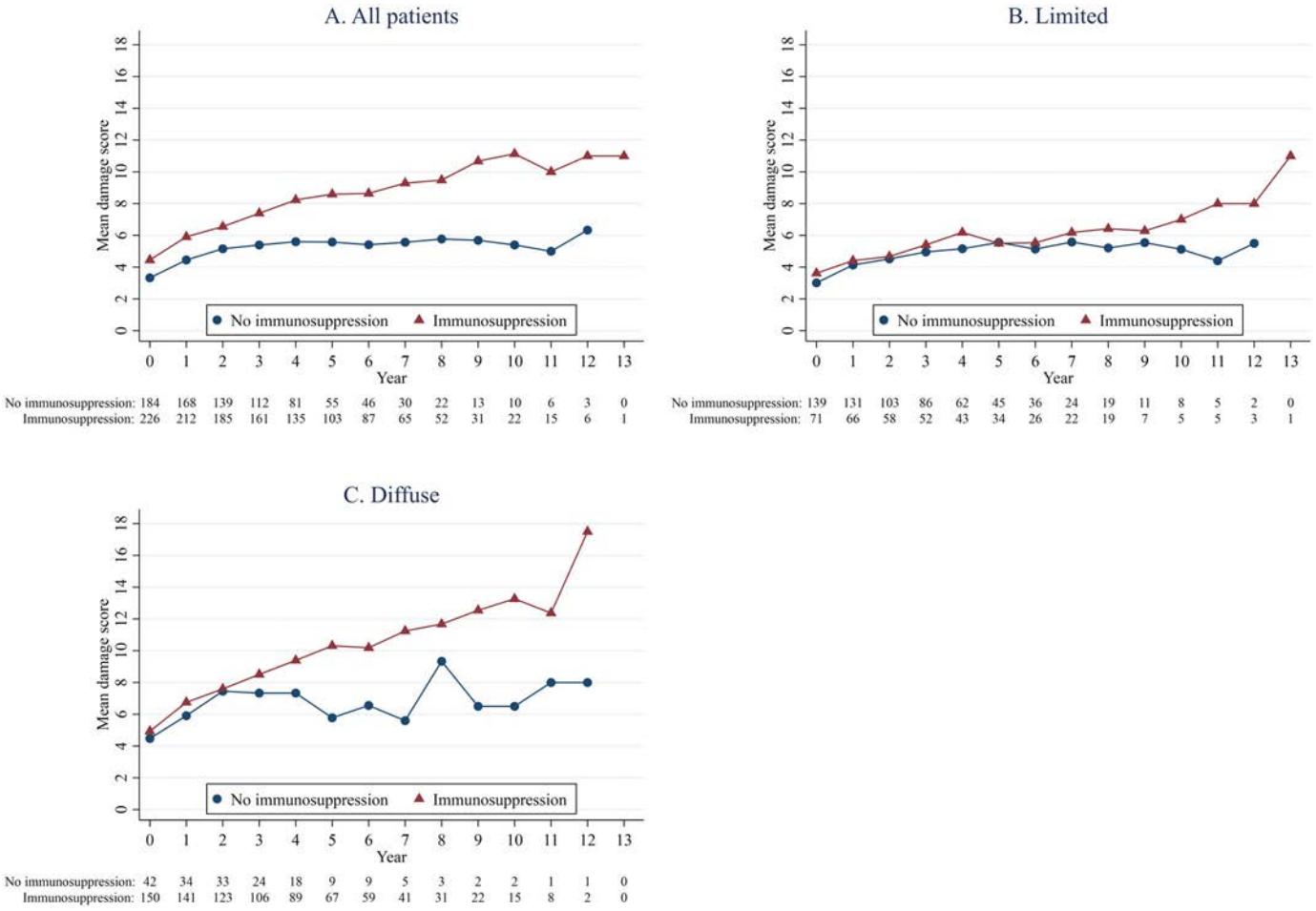


Figure 1. Mean Scleroderma Clinical Trials Consortium Damage Index scores. (A) All patients. (B) Patients with limited cutaneous systemic sclerosis. (C) Patients with diffuse cutaneous systemic sclerosis. All data are from before propensity weighting.

in Supplementary Table 3. The balancing of sex, mRSS, arthritis, and immunosuppression before baseline were outside expected values, but balancing was good for other variables and adequate for FVC. Table 4 shows the association between exposure to immunosuppression and the change of damage scores adjusted for possible confounders. Exposure to ISs at

Table 2. Marginal structural GEE model using IPTW to assess the association between exposure to immunosuppression and the change of damage scores, adjusted for potential confounders in patients with limited cutaneous scleroderma*

Characteristic	Without IPTW		With IPTW	
	beta (95% CI)	*P* value	beta (95% CI)	*P* value
Exposure to immunosuppressive drugs	−0.22 (−0.40 to −0.04)	0.019	−0.34 (−0.59 to −0.09)	0.007
Female	−0.16 (−0.40 to 0.07)	0.177	−0.16 (−0.37 to 0.05)	0.144
Age, y	0.004 (−0.004 to 0.01)	0.319	0.001 (−0.01 to 0.01)	0.774
Disease duration	−0.07 (−0.27 to 0.13)	0.513	0.09 (−0.12 to 0.30)	0.407
ACA	−0.03 (−0.32 to 0.25)	0.821	0.01 (−0.26 to 0.27)	0.969
ATA	−0.08 (−0.29 to 0.13)	0.452	−0.13 (−0.34 to 0.08)	0.219
Immunosuppressive drugs before baseline	0.38 (−0.24 to 0.99)	0.230	0.20 (−0.71 to 1.11)	0.661
Damage score at baseline	−0.03 (−0.06 to 0.01)	0.170	−0.002 (−0.04 to 0.04)	0.992
FVC predicted, %	−0.01 (−0.01 to −0.002)	0.008	−0.01 (−0.01 to −0.001)	0.017
CSRG vs ASIG	0.13 (−0.03 to 0.31)	0.118	0.02 (−0.11 to 0.16)	0.738

* ACA, anti-centromere antibody; ASIG, Australian Scleroderma Interest Group; ATA, anti-topoisomerase antibody; CI, confidence interval; CSRG, Canadian Scleroderma Research Group; FVC, forced vital capacity; GEE, generalized estimating equation; IPTW, inverse probability of treatment weighting.

Table 3. Baseline characteristics of diffuse cutaneous scleroderma patients (dcSSc) stratified according to exposure to immunosuppression at baseline or during follow–up (*N* = 192)*

	Ever exposed at any visit (*n* = 150)		Never exposed at any visits (*n* = 142)		*P* values
		Missing values *N*		Missing values *N*	
Age, years (mean ± SD)	50.0 ± 12.1		55.5 ± 10.7		0.008
Female, *N* (%)	110 (73.3%)		32 (76.2%)		0.709
Caucasian, *N* (%)	122 (81.3%)		38 (90.5%)		0.160
Education (>high school), *N* (%)	69 (48.6%)	8	18 (42.9%)	0	0.513
Smoking in the past or currently, N (%)	81 (54.4%)	1	29 (69.0%)	0	0.089
Disease duration, years (mean ± SD)	1.1 ± 0.5		1.1 ± 0.5		0.723
mRSS (0–51) (media, IQR)	23.5 (15–29)	2	20 (16–24)	1	0.296
Interstitial lung disease, N (%)	53 (35.3%)		15 (35.7%)		0.964
FVC, % predicted (mean ± SD)	88.3 ± 17.4	1	90.1 ± 23.3	0	0.585
DLCO, % predicted (mean ± SD)	71.2 ± 20.7	11	73.0 ± 16.1	8	0.639
Tendon friction rubs, *N* (%)	41 (27.3%)		11 (26.2%)		0.883
Inflammatory arthritis, *N* (%)	53 (36.3%)	4	10 (25.0%)	2	0.181
Autoantibodies					
Anti-centromere, *N* (%)	4 (3.0%)	15	7 (18.0%)	3	0.001
Anti-topoisomerase, *N* (%)	44 (33.1%)	17	4 (10.3%)	3	0.005
Anti-RNA polymerase III, *N* (%)	55 (44.0%)	25	16 (44.0%)	6	0.962
C–reactive protein, mg/L (median, IQR)	5 (3–12)	10	4.4 (1.9–9.2)	2	0.110
Damage score at baseline (median, IQR)	5 (2–7)		3 (1–6)		1.02
CSRG patients, *N* (%)	77 (51.3%)		26 (61.9%)		0.225
Immunosuppressants prior to baseline, *N* (%)	14 (9.7%)		6 (12.8%)		0.585

* CSRG, Canadian Scleroderma Research Group; DLCO, diffusing capacity of carbon monoxide; FVC, forced vital capacity; IS, immunosuppressives; mRSS, modified Rodnan skin score.

any time was not associated with damage scores, but exposure to ISs before baseline and lower FVC values were associated with more damage accrual with IPTW.

DISCUSSION

To our knowledge, this is the first study that has examined the effect of receiving ISs and damage accrual in patients with

Table 4. Marginal structural GEE model using IPTW to assess the association between exposure to immunosuppression and the change of damage scores, adjusted for potential confounders in patients with diffuse cutaneous scleroderma*

Characteristic	Without IPTW		With IPTW	
	β (95% CI)	*P* value	β (95% CI)	*P* value
Exposure to immunosuppressive drugs	−0.07 (−0.36 to 0.23)	0.667	−0.08 (−0.41 to 0.25)	0.627
Female	−0.34 (−0.66 to −0.02)	0.040	−0.31 (−0.72 to −0.09)	0.132
Age, y	0.001 (−0.01 to 0.01)	0.911	0.004 (−0.01 to 0.02)	0.525
Disease duration, y	−0.24 (−0.57 to 0.08)	0.144	−0.24 (−0.62 to 0.14)	0.210
ATA	0.003 (−0.30 to 0.31)	0.987	−0.04 (−0.38 to 0.30)	0.826
RNAP	0.13 (−0.20 to 0.46)	0.433	0.01 (−0.38 to 0.40)	0.969
Exposure to immunosuppressive drugs before baseline	0.20 (−0.05 to 0.45)	0.109	0.25 (0.02–0.47)	0.036
Damage score at baseline	−0.01 (−0.05 to 0.02)	0.506	−0.01 (−0.04 to 0.02)	0.449
FVC predicted, %	−0.01 (−0.02 to −0.003)	0.004	−0.01 (−0.02 to −0.002)	0.010
CSRG vs ASIG	0.13 (−0.02 to 0.36)	0.238	0.17 (−0.12 to 0.46)	0.255

* ASIG, Australian Scleroderma Interest Group; ATA, anti-topoisomerase antibody; CI, confidence interval; CSRG, Canadian Scleroderma Research Group; FVC, forced vital capacity; GEE, generalized estimating equation; IPTW, inverse probability of treatment weighting; RNAP, anti–RNA polymerase III antibody.

SSc using the recently developed SCTC-DI. Because the cohorts studied were observational, IPTW was used to account for the different aspects of disease that may have encouraged the prescription of ISs and to remove as much bias as possible that these different disease characteristics may have had on the outcome. Because damage accrual is so different in patients with lcSSc and dcSSc, and use of ISs was more frequent in patients with dcSSc, we examined each cutaneous subset of patients separately. As such, we were able to find a positive effect of prevention of damage accrual in the subgroup with lcSSc but not in the subgroup with dcSSc.

IS agents have been prescribed extensively to patients with SSc.[2,16,17,28] Many recommendations and guidelines suggest prescribing ISs for patients with skin and lung disease,[29–34] and in practice, many scleroderma experts agree and follow these recommendations.[35,36] Although fibrosis may be the proximate cause of most damage in SSc, there has been some evidence that receiving ISs may prevent fibrosis and thus prevent accrual of damage.[10,11,17,29,37–43] Most such studies, however, have concentrated on damage to individual organ systems such as the lung or skin, but the development of the SCTC-DI has allowed us to determine the effect of receiving ISs on a more global measure of damage.[12–14]

There are some important limitations to our study. The SCTC-DI items were determined to a large extent by expert consensus and weighted against mortality and morbidity. We were aware at the time of creating the SCTC-DI that not all existing databases would include each item, and we agree that some of the items missing from our databases such as small-joint contractures are important. Unfortunately, we cannot objectively determine how this might have affected our results.

Any study of the effect of an exposure on an outcome using data from observational cohorts is affected by bias. Propensity scoring, such as IPTW, is a method that allows an investigator to assess the impact of an exposure of interest using observational cohorts, in which randomization is not performed, instead of in a prospective randomized trial. Certain patient characteristics that are a common cause of both the observed exposure and the outcome may confound the relationship under study, leading to an over- or underestimation of the true effect. IPTW attempts to correct for these when studying the effect of an exposure in an observational cohort.[44] Somewhat to our surprise, and only partially in keeping with our hypothesis that IS use would prevent damage accrual, we found that IS use prevented damage accrual only in the subset with lcSSc. It is important therefore to try to understand whether the absence of an effect in patients with dcSSc was a true finding or if IPTW was not sufficient to adjust for the biases inherent in using observational cohort data.

We considered several possibilities that may have prevented us from finding a true effect on damage accrual in patients with dcSSc. One is that patients with dcSSc already have considerably more damage within the first two years of disease than those with lcSSc, and therefore, the chance of developing more damage might seem to be less, and thus, it would be more difficult to show prevention of damage accrual. We have found, however, that although the damage index is higher in patients with dcSSc than in those with lcSSc very early in disease progression, in fact, patients with dcSSc continue to accrue damage over subsequent years at a faster rate than patients with lcSSc.[12,13] Perhaps ISs are just not adequate to suppress this faster rate of damage accrual.

It is also possible that the weighting itself in IPTW was not adequate to account for some of the differences between the populations. Assessing the adequacy of the weighting is done by assessing the standardized differences between groups for all baseline characteristics both before and after weighting. Supplementary Tables 2 and 3 compare the difference in means between groups in units of SD. A standardized difference of <10% (<0.1) may be considered a negligible imbalance between groups.[26] From Supplementary Table 2 use before baseline. However, from Supplementary Table 3, for dcSSc, weighting was not very good for sex, anti-centromere antibodies, mRSS, and arthritis. This may have contributed to our inability to find an effect of ISs in the group with dcSSc.

As pointed out by Chesnaye et al,[44] by accounting for any differences in measured baseline characteristics, the propensity score aims to approximate what would have been achieved through randomization in a randomized controlled trial (RCT); ie, pseudorandomization. In contrast to true randomization, the propensity score can only account for measured confounders, not for any unmeasured confounders.[44] Therefore, an imbalance in some unmeasured confounder may also have played a role in our inability to find an effect of ISs in patients with dcSSc. This may have been true in the current patients with dcSSc, as indicated by the observation that certain characteristics such as more frequent ILD and arthritis, which might be expected to be associated with more use of ISs, were indeed more prevalent in the group of patients with lcSSc who were exposed to ISs but were found in similar proportions in the groups with dcSSc who were and were not exposed to ISs. This suggests that the decision to prescribe ISs to the patients with dcSSc may have been related to an unmeasured confounder in the subset with dcSSc.

Our study also has important strengths. By combining two observational cohorts, we were able to find a large enough number of patients in both subsets with cutaneous SSc to assess our hypotheses. We have also used a now-accepted technique of studying the effect of an exposure to a treatment in an observational cohort and have found that at least it seems to be effective in one of the subsets with cutaneous SSc. With these caveats in mind, and with our findings of a protective effect of receiving ISs on damage accrual in patients with lcSSc, we feel that the effect of ISs on damage in patients with dcSSc has not yet been ruled out. We anticipate the use of the SCTC-DI in future RCTs in SSc and hope that those trials will be able to answer more definitively

the question of the benefit of ISs in patients with dcSSc for the prevention of damage accrual.

ACKNOWLEDGMENT

We thank Daniela Solomon for help in submitting the paper.

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 Baron 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. Maher TM, Tudor VA, Saunders P, et al. Rituximab versus intravenous cyclophosphamide in patients with connective tissue disease-associated interstitial lung disease in the UK (RECITAL): a double-blind, double-dummy, randomised, controlled, phase 2b trial. Lancet Respir Med 2023;11:45–54.
2. Thoreau B, Chaigne B, Renaud A, et al. Treatment of systemic sclerosis. Presse Med 2021;50:104088.
3. Volkmann ER, Tashkin DP, Sim M, et al. Cyclophosphamide for systemic sclerosis-related interstitial lung disease: a comparison of Scleroderma Lung Study I and II. J Rheumatol 2019;46:1316–1325.
4. Gordon JK, Martyanov V, Franks JM, et al. Belimumab for the treatment of early diffuse systemic sclerosis: results of a randomized, double-blind, placebo-controlled, pilot trial. Arthritis Rheumatol 2018;70:308–316.
5. Goldin JG, Kim GHJ, Tseng CH, et al. Longitudinal changes in quantitative interstitial lung disease on computed tomography after immunosuppression in the Scleroderma Lung Study II. Ann Am Thorac Soc 2018;15:1286–1295.
6. Wanchu A, Suryanaryana BS, Sharma S, et al. High-dose prednisolone and bolus cyclophosphamide in interstitial lung disease associated with systemic sclerosis: a prospective open study. Int J Rheum Dis 2009;12:239–242.
7. Várai G, Earle L, Jimenez SA, et al. A pilot study of intermittent intravenous cyclophosphamide for the treatment of systemic sclerosis associated lung disease. J Rheumatol 1998;25:1325–1329.
8. Karalilova RV, Batalov ZA, Sapundzhieva TL, et al. Tofacitinib in the treatment of skin and musculoskeletal involvement in patients with systemic sclerosis, evaluated by ultrasound. Rheumatol Int 2021;41:1743–1753.
9. Hachulla E, Agard C, Allanore Y, et al. French recommendations for the management of systemic sclerosis. Orphanet J Rare Dis 2021;16:322.
10. Bruni C, Tashkin DP, Steen V, et al. Intravenous versus oral cyclophosphamide for lung and/or skin fibrosis in systemic sclerosis: an indirect comparison from EUSTAR and randomised controlled trials. Clin Exp Rheumatol 2020;38(Suppl 125):161–168.
11. Burt RK, Shah SJ, Dill K, et al. Autologous non-myeloablative haemopoietic stem-cell transplantation compared with pulse cyclophosphamide once per month for systemic sclerosis (ASSIST): an open-label, randomised phase 2 trial. Lancet 2011;378:498–506.
12. Baron M, Barbacki A, Man A, et al. Prediction of damage trajectories in systemic sclerosis using group-based trajectory modeling. Rheumatology (Oxford) 2023;62:3059–3066.
13. Barbacki A, Baron M, Wang M, et al. Damage trajectories in systemic sclerosis using group-based trajectory modeling. Arthritis Care Res (Hoboken) 2023;75:640–647.
14. Ferdowsi N, Huq M, Stevens W, et al. Development and validation of the Scleroderma Clinical Trials Consortium Damage Index (SCTC-DI): a novel instrument to quantify organ damage in systemic sclerosis. Ann Rheum Dis 2019;78:807–816.
15. Tay T, Stevens W, Baron M, et al. Early accrual of organ damage in scleroderma: rationale for derivation and validation of a disease damage index in systemic sclerosis. Clin Exp Rheumatol 2014;81:S6.
16. Sierra-Sepúlveda A, Esquinca-González A, Benavides-Suárez SA, et al. Systemic sclerosis pathogenesis and emerging therapies, beyond the fibroblast. Biomed Res Int 2019;2019:4569826.
17. Verrecchia F, Wang Y, Vija L, et al. Evidence of an antifibrotic effect of immunosuppressive drugs: applications in the treatment of systemic sclerosis. Expert Rev Clin Immunol 2009;5:35–43.
18. Goldin J, Elashoff R, Kim HJ, et al. Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. Chest 2009;136:1333–1340.
19. Denton CP, Black CM, Abraham DJ. Mechanisms and consequences of fibrosis in systemic sclerosis. Nat Clin Pract Rheumatol 2006;2:134–144.
20. Hasegawa M, Fujimoto M, Takehara K, et al. Pathogenesis of systemic sclerosis: altered B cell function is the key linking systemic autoimmunity and tissue fibrosis. J Dermatol Sci 2005;39:1–7.
21. 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. Ann Rheum Dis 2013;72:1747–1755.
22. Ferdowsi N, Huq M, Stevens W, et al. Development and validation of the Scleroderma Clinical Trials Consortium Damage Index (SCTC-DI): a novel instrument to quantify organ damage in systemic sclerosis. Ann Rheum Dis 2019;78:807–816.
23. Steele R, Hudson M, Lo E, et al; Canadian Scleroderma Research Group. Clinical decision rule to predict the presence of interstitial lung disease in systemic sclerosis. Arthritis Care Res (Hoboken) 2012;64:519–524.
24. Hsu VM, Moreyra AE, Wilson AC, et al. Assessment of pulmonary arterial hypertension in patients with systemic sclerosis: comparison of noninvasive tests with results of right-heart catheterization. J Rheumatol 2008;35:458–465.
25. He H, Hu J, He J. Overview of propensity score methods. In: He H, Wu P, Chen DG, eds. Statistical Causal Inferences and Their Applications in Public Health Research. Springer International Publishing; 2016:29–48.
26. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. J Thorac Cardiovasc Surg 2007;134:1128–1135.
27. 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:S84–S90.e1.
28. Boleto G, Avouac J, Allanore Y. An update on recent randomized clinical trials in systemic sclerosis. Joint Bone Spine 2021;88:105184.
29. Zhao M, Wu J, Wu H, et al. Clinical treatment options in scleroderma: recommendations and comprehensive review. Clin Rev Allergy Immunol 2022;62:273–291.

30. Kowal-Bielecka O, Fransen J, Avouac J, et al. Update of EULAR recommendations for the treatment of systemic sclerosis. Ann Rheum Dis 2017;76:1327–1339.
31. Rahaghi FF, Hsu VM, Kaner RJ, et al. Expert consensus on the management of systemic sclerosis-associated interstitial lung disease. Respir Res 2023;24:6.
32. Kowal-Bielecka O, Landewé R, Avouac J, et al. EULAR recommendations for the treatment of systemic sclerosis: a report from the EULAR Scleroderma Trials and Research group (EUSTAR). Ann Rheum Dis 2009;68:620–628.
33. Avouac J, Kowal-Bielecka O, Landewe R, et al. European League Against Rheumatism (EULAR) Scleroderma Trial and Research group (EUSTAR) recommendations for the treatment of systemic sclerosis: methods of elaboration and results of systematic literature research. Ann Rheum Dis 2009;68:629–634.
34. Asano Y, Jinnin M, Kawaguchi Y, et al. Diagnostic criteria, severity classification and guidelines of systemic sclerosis. J Dermatol 2018; 45:633–691.
35. Walker KM, Pope J. Treatment of systemic sclerosis complications: what to use when first-line treatment fails–a consensus of systemic sclerosis experts. Semin Arthritis Rheum 2012;42:42–55.
36. de Vries-Bouwstra JK, Allanore Y, Matucci-Cerinic M, et al. Worldwide expert agreement on updated recommendations for the treatment of systemic sclerosis. J Rheumatol 2020;47:249–254.
37. Mendoza FA, Nagle SJ, Lee JB, et al. A prospective observational study of mycophenolate mofetil treatment in progressive diffuse cutaneous systemic sclerosis of recent onset. J Rheumatol 2012;39: 1241–1247.
38. Kim HJ, Brown MS, Elashoff R, et al. Quantitative texture-based assessment of one-year changes in fibrotic reticular patterns on HRCT in scleroderma lung disease treated with oral cyclophosphamide. Eur Radiol 2011;21:2455–2465.
39. McSweeney PA, Nash RA, Sullivan KM, et al. High-dose immunosuppressive therapy for severe systemic sclerosis: initial outcomes. Blood 2002;100:1602–1610.
40. Au K, Mayes MD, Maranian P, et al. Course of dermal ulcers and musculoskeletal involvement in systemic sclerosis patients in the scleroderma lung study. Arthritis Care Res (Hoboken) 2010;62:1772–1778.
41. Hoyles RK, Ellis RW, Wellsbury J, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. Arthritis Rheum 2006;54:3962–3970.
42. Namas R, Tashkin DP, Furst DE, et al. Efficacy of mycophenolate mofetil and oral cyclophosphamide on skin thickness: post hoc analyses from two randomized placebo-controlled trials. Arthritis Care Res (Hoboken) 2018;70:439–444.
43. Saunders P, Tsipouri V, Keir GJ, et al. Rituximab versus cyclophosphamide for the treatment of connective tissue disease-associated interstitial lung disease (RECITAL): study protocol for a randomised controlled trial. Trials 2017;18:275.
44. Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. Clin Kidney J 2022;15:14–20.

APPENDIX A: AUSTRALIAN SCLERODERMA INTEREST GROUP

The following are members of the Australian Scleroderma Interest Group who were not named as authors but all of whom collected data for this paper and reviewed the manuscript: Joanne Sahhar, Monash University; Nava Ferdowsi, Victoria, Australia; Kathleen Morrisroe, University of Melbourne; Laura Ross, University of Melbourne; Gene Siew Ngian, University of Melbourne; Jennifer Walker, Flinders University; Janet Roddy, Fiona Stanely Hospital; and Lauren Host, Sir Charles Gairdner Hospital.

APPENDIX B: CANADIAN SCLERODERMA RESEARCH GROUP

The following are members of the Canadian Scleroderma Research Group who were not named as authors but all of whom collected data for this paper and reviewed the manuscript: Mohammed Osman, University of Alberta; Peter Docherty, Moncton Hospital; Paul Fortin, Universite Laval; Marvin Jacob Fritzler, University of Calgary; Genevieve Gyger, McGill University; Alena Ikic, Universite Laval; Niall Jones, University of Alberta; Elzbieta Kaminska, University of Calgary; Maggie Larche, McMaster University; Sophie Ligier, Hopital Maisonneuve-Rosemont; Ada Man, University of Manitoba; Ariel Masetto, University of Sherbrooke; Janet Pope, Western University; David Robinson, University of Manitoba; and Tatiana Sofia Rodriguez-Reyna, Instituto Nacional de Ciencias Médicas y Nutrición.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 649–657
DOI 10.1002/acr.25469

American College of Rheumatology
Empowering Rheumatology Professionals

Oral Glucocorticoids for Skin Fibrosis in Early Diffuse Systemic Sclerosis: A Target Trial Emulation Study From the European Scleroderma Trials and Research Group Database

Denis Mongin,[1] Marco Matucci-Cerinic,[2] Ulrich A. Walker,[3] Oliver Distler,[4] Radim Becvar,[5] Elise Siegert,[6] Lidia P. Ananyeva,[7] Vanessa Smith,[8] Juan Jose Alegre-Sancho,[9] Sule Yavuz,[10] Massimiliano Limonta,[11] Gabriela Riemekasten,[12] Elena Rezus,[13] Madelon Vonk,[14] Marie-Elise Truchetet,[15] Francesco Del Galdo,[16] Delphine S. Courvoisier,[1] and Michele Iudici,[1] on behalf of the EUSTAR Collaborators

Objective. The objective of this study is to evaluate whether adding oral glucocorticoids to immunosuppressive therapy improves skin scores and ensures safety in patients with early diffuse cutaneous systemic sclerosis (dcSSc).

Methods. We performed an emulated randomized trial comparing the changes from baseline to 12 ± 3 months of the modified Rodnan skin score (mRSS: primary outcome) in patients with early dcSSc receiving either oral glucocorticoids (≤20 mg/day prednisone equivalent) combined with immunosuppression (treated) or immunosuppression alone (controls), using data from the European Scleroderma Trials and Research Group. Secondary end points were the difference occurrence of progressive skin or lung fibrosis and scleroderma renal crisis. Matching propensity score was used to adjust for baseline imbalance between groups.

Results. We matched 208 patients (mean age 49 years; 33% male; 59% anti-Scl70), 104 in each treatment group, obtaining comparable characteristics at baseline. In the treated group, patients received a median prednisone dose of 5 mg/day. Mean mRSS change at 12 ± 3 months was similar in the two groups (decrease of 2.7 [95% confidence interval {95% CI} 1.4–4.0] in treated vs 3.1 [95% CI 1.9–4.4] in control, $P = 0.64$). Similar results were observed in patients with shorter disease duration (≤ 24 months) or with mRSS ≤22. There was no between-group difference for all prespecified secondary outcomes. A case of scleroderma renal crisis occurred in both groups.

Conclusion. We did not find any significant benefit of adding low-dose oral glucocorticoids to immunosuppression for skin fibrosis, and at this dosage, glucocorticoid did not increase the risk of scleroderma renal crisis.

INTRODUCTION

Systemic sclerosis (SSc) is a complex, multisystemic autoimmune disease characterized by high mortality and morbidity.[1] The two main recognized disease subsets, ie, limited cutaneous (lc) and diffuse cutaneous (dc) SSc,[2] present different evolution and prognosis. Patients with dcSSc manifest fibrotic complications and decline in organ function early during the disease course, resulting in poor quality of life, and leading to a mortality rate that is five to eight times higher than that observed in the general population.[3] Currently, available treatment offers limited benefit, and hematopoietic stem cell transplantation should be

Supported by the Project Recherche & développement from Geneva University Hospitals.

[1]Denis Mongin, PhD, Delphine S. Courvoisier, PhD, Michele Iudici, MD, PhD, MPH: Geneva University Hospitals, Geneva, Switzerland; [2]Marco Matucci-Cerinic, MD: IRCCS San Raffaele Scientific Institute, IRCCS San Raffaele Hospital, and Vita-Salute San Raffaele University, Milan, Italy; [3]Ulrich A. Walker, MD: University Hospital Basel, Basel, Switzerland; [4]Oliver Distler, MD: University Hospital Zurich, Zurich, Switzerland; [5]Radim Becvar, MD: University, Prague, Czech Republic; [6]Elise Siegert, MD: Charité University Hospital, Berlin, Germany; [7]Lidia P. Ananyeva, MD: V.A. Nasonova Research Institute of Rheumatology Russian Federation, Moskow, Russia; [8]Vanessa Smith, MD, PhD: Ghent University, Ghent University Hospital, and VIB Inflammation Research Center, Ghent, Belgium; [9]Juan Jose Alegre-Sancho, MD: Hospital Universitario Doctor Peset, Valencia, Spain; [10]Sule Yavuz, MD: Istanbul Bilim University, Altunizade-Istanbul, Turkey; [11]Massimiliano Limonta, MD: ASST Papa Giovanni XXIII, Begamo, Italy; [12]Gabriela Riemekasten, MD: Klinik Für Rheumatologie Und Klinische Immunologie, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany; [13]Elena Rezus, MD: Grigore T. Popa University of Medicine and Pharmacy Iasi, Rehabilitation Hospital Iasi, Romania; [14]Madelon Vonk, MD: Radboud University Medical Center, Nijmegen, The Netherlands; [15]Marie-Elise Truchetet, MD, PhD: National Reference Center for Systemic Autoimmune Rare Diseases, Bordeaux University Hospital, Hôpital Pellegrin, Bordeaux, France; [16]Francesco Del Galdo, MD, PhD: Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, United Kingdom.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25469).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25469.

Address correspondence via email to Michele Iudici, MD, PhD, MPH, at michele.iudici@hcuge.ch.

Submitted for publication July 10, 2024; accepted in revised form November 4, 2024.

SIGNIFICANCE & INNOVATIONS

- The effect of oral glucocorticoids in addition to immunosuppression on skin fibrosis progression in patients with early diffuse cutaneous systemic sclerosis is not known.
- This study managed to emulate a randomized control trial of 208 patients with early diffuse cutaneous systemic sclerosis, of whom 104 were treated with oral glucocorticoids and immunosuppression.
- Our results suggest that the use of oral glucocorticoids is not needed to slow down skin fibrosis in patients with early diffuse cutaneous systemic sclerosis.

considered in selected cases.[4] Guidelines recommend that immunosuppression (mycophenolate mofetil, rituximab, cyclophosphamide) should be promptly introduced to control symptoms and prevent fibrotic complications.[5–7]

Whether adding oral glucocorticoids to immunosuppression is beneficial to slowing down skin fibrosis progression in patients with early dcSSc is a matter of debate.[8–10] In past years, research has failed to define the exact role of these compounds for this indication, mainly because of limitations in study design (nonrandomized studies,[11] small randomized studies at high risk of bias[10,12]). Recently, a randomized controlled trial investigating the impact of adding oral glucocorticoids to routine care on skin fibrosis progression and quality of life ended prematurely for insufficient patient accrual due to COVID-19 pandemic.[13] In the absence of clear evidence, glucocorticoids are being widely prescribed to patients with dcSSc, and often for long periods,[8,9,14,15] with the consequent risk of drug toxicity and treatment-related damage accrual.

In the context of a rare disease in which conducting a randomized controlled trial is difficult and may not succeed, comparative effectiveness studies on large samples of routinely collected data can provide guidance for management. Target trial emulation can help reduce bias in the effect estimates derived from observational analyses.[16] This study aims to evaluate whether adding oral glucocorticoids to immunosuppression for skin fibrosis improved skin scores at one year and was safe in terms of risk of scleroderma renal crisis.

METHODS

Ethics and regulations. This study was conducted in compliance with the Declaration of Helsinki. Local ethic committee permission for each European Scleroderma Trials and Research Group (EUSTAR) center and informed consent, where appropriate according to local ethic regulations, was obtained before EUSTAR enrollment. For a list of the EUSTAR collaborators, please see Appendix A.

Specification of the target trial. This observational analysis was designed to emulate a target trial (ie, a hypothetical pragmatic trial that would have answered the causal question of interest[17]) on the use of immunosuppression plus oral glucocorticoids versus immunosuppression alone to improve skin fibrosis in patients with early dcSSc (see eligibility criteria below). The main outcome was the between-arm difference in skin fibrosis change at one year from baseline.

EUSTAR database. This analysis was based on data from patients with SSc enrolled in EUSTAR.[1,18] This international database prospectively collects data from patients with SSc seen in routine care at least annually from >200 centers from all the continents. For inclusion, patients have to meet the 2013 American College of Rheumatology/EULAR classification criteria for SSc.[19]

Eligibility criteria. We included patients with SSc aged 18 years or more, presenting with a dcSSc subset according to Leroy et al,[2] a modified Rodnan skin score (mRSS) ≥7, with ≤5 years disease duration since the first non-Raynaud symptom, who had started either immunosuppression monotherapy or immunosuppression plus oral glucocorticoids (≤20 mg of prednisone equivalent per day) at inclusion between January 2009 and December 2020. The criterion of mRSS ≥7 equals the minimal score that is required to qualify as dcSSc and was deliberately chosen not to miss early patients who will become progressive later on. We excluded patients with a previous renal crisis, patients receiving daily >20 mg prednisone equivalent at inclusion, patients receiving or who had received pulse methylprednisolone within six weeks from the inclusion, patients with an associated connective tissue disease, and patients who had received hematopoietic stem cell transplantation or lung or heart transplantation.

Interventions. We compared the following two treatment strategies at inclusion: 1) immunosuppression alone. Patients in this group could have received in the timeframe considered intravenous or oral cyclophosphamide, methotrexate, mycophenolate mofetil, azathioprine, rituximab, and/or tocilizumab without concomitant glucocorticoids ("control group"); 2) immunosuppression (as defined above) plus oral glucocorticoids at a daily dose ≤20 mg prednisone equivalent ("treated group").

Primary and secondary outcomes. The primary outcome measure was the between-group mean difference of mRSS at 12 ± 3 months from baseline. EUSTAR centers are advised that the same investigator is performing the mRSS on follow-up visits in individual patients, and EUSTAR investigators are trained on a regular basis in how to perform the mRSS.[20,21] The follow-up time of 12 months was selected since it is considered a relevant timeframe to detect significant changes in mRSS, and is therefore used in many clinical

studies.[22,23] Secondary end points, evaluated at 12 ± 3 months from baseline, were as follows:

- development of progressive skin fibrosis (increase in mRSS of 5 points and ≥25% from baseline);
- development of progressive lung fibrosis (either decrease in either forced vital capacity [FVC] ≥10% or both decrease of FVC ≥5% and decrease in diffusing capacity of the lung for carbon monoxide ≥15%);
- development of progressive skin fibrosis and development of progressive lung fibrosis[21,24,25];
- incidence of scleroderma renal crisis.

Subgroup analyses. We performed two subgroup analyses for the primary outcome. The first focused on patients with a disease duration from non-Raynaud onset ≤24 months, who are those with more active inflammation and who are therefore more likely to respond to glucocorticoids.[21,26] These patients are also more likely to have worsening skin and lung disease.[27] The second included the subset of patients with an mRSS value <22 in order to enrich the sample of patients with the highest likelihood to have a progressive skin disease.[21] Imputation and matching of the groups were re-performed for each subgroup analysis (see next section).

Statistical analyses. Two sets of analysis were performed for each outcome; the main one included patients having at least a measured outcome difference at 12 ± 3 months. The second one was a sensitivity analysis considering all the patients fulfilling the eligibility criteria. In this analysis, we addressed missing data for the main outcome by incorporating them along with the covariates during the imputation process for missing variables. For both sets of analysis, treatment groups were first matched using propensity score matching with a 1:1 ratio using nearest neighbor

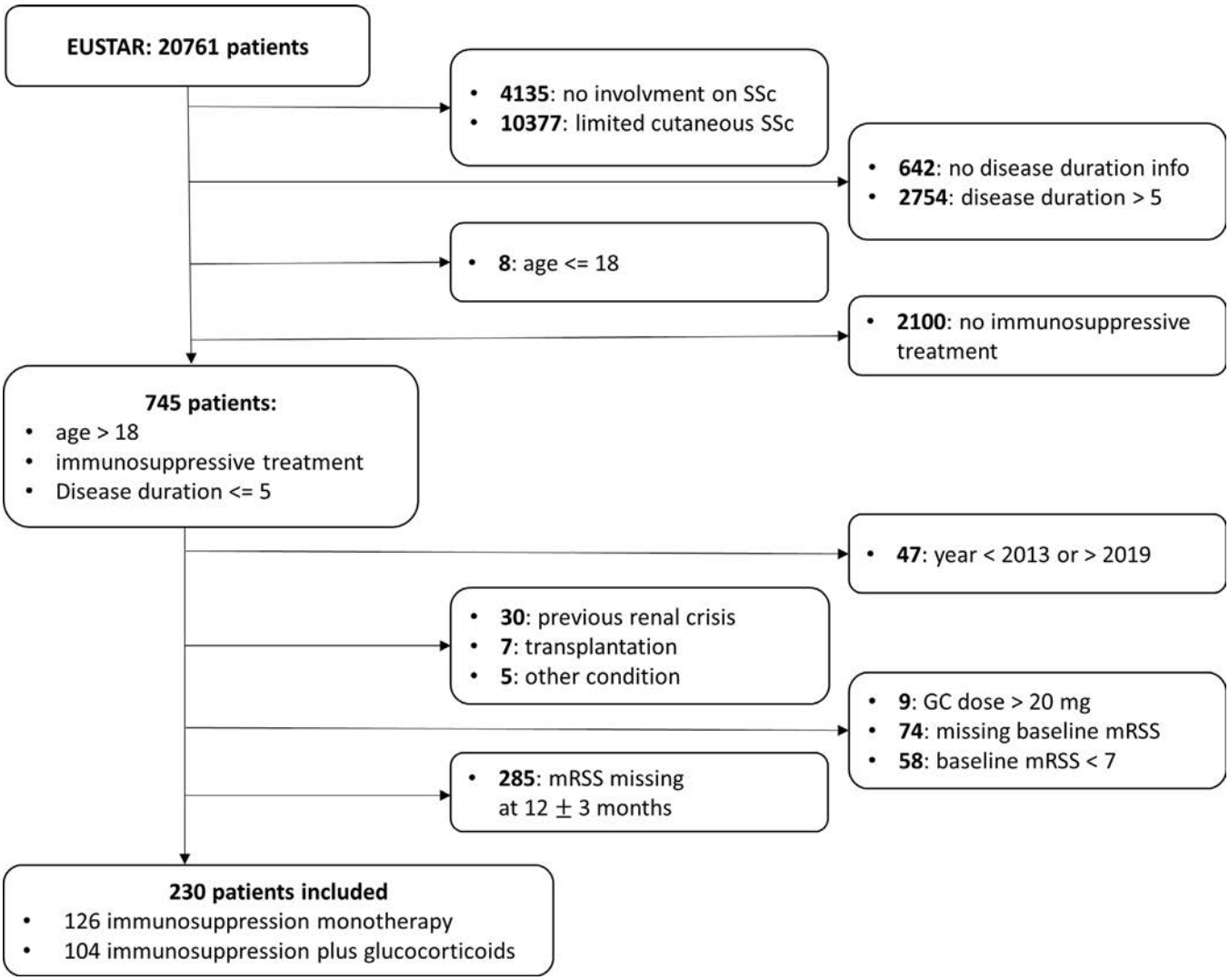


Figure 1. Flowchart of the selection process of the target population from the EUSTAR database. EUSTAR, European Scleroderma Trials and Research Group; GC, glucocorticoid; mRSS, modified Rodnan skin score; SSc, systemic sclerosis.

matching. The propensity score was assessed using logistic multivariable regression with use of glucocorticoids as dependent variable and a set of covariates as independent variables. These were identified by the analysis of the literature on the topic[21,26–29] 29 and discussion among some of the authors (DM, FDG, DSC, and MI). Priority was given to prognostic variables, whereas variables strongly associated with treatment but not—or weakly—with the outcome (instrumental variables) can induce unstable weights with little or no gain in terms of bias reduction. We included in the model age, sex, disease duration (from the first non-Raynaud onset), baseline mRSS, presence of joint synovitis, immunosuppression treatment type (cyclophosphamide, methotrexate, mycophenolate mofetil, rituximab, other), anti-topoisomerase and anti-RNA polymerase III status, presence of tendon friction rubs, smoking status, and year of enrollment. Missing covariates were handled using multiple imputation with chained equations, with 50 samples and 10 iterations, considering all covariates and the outcome in the imputation model. The predicted propensity score was the mean of the predicted scores of each sample.

Descriptive statistics used paired *t*-test for continuous variables and McNemar or Friedman tests for dichotomous variables to account for the matching of the groups. The difference in outcomes between the two matched groups was analyzed using linear regression predicting the main outcome or logistic regression predicting the secondary outcomes as a function of the baseline use of glucocorticoids and with a random intercept on the matched patients to account for the matching of the groups.

For the sensitivity analysis, matching was performed the same way. The outcome was imputed considering the outcome and all covariates in the imputation model. The estimation of the effect of baseline use of glucocorticoids was then pooled from regression estimates following Rubin's rules.[30] The analyses were computed using R software version 4.0.3,[31] with mice library for the multiple imputation.[32] Data are available on reasonable request.

RESULTS

Study population. Among 20,761 patients with SSc from the EUSTAR database assessed for eligibility, 745 had at least 1 visit with the information proving that they had a disease duration <5 years, received immunosuppression, and were adult patients. An additional 515 patients did not meet inclusion criteria (see Figure 1). Of the 230 patients who met the eligibility criteria (see Supplementary Table 1 for key characteristics), we successfully performed 1:1 matching for 208 patients; 104 received immunosuppression without glucocorticoids, and 104 received a combination of immunosuppression and glucocorticoids. The two groups of patients had similar baseline characteristics (Table 1): 67% were female, with a mean age of 49 years, a mean disease duration of 2.5 years, and a median mRSS of 18, with 43% of the patients having interstitial lung disease identified by lung x-ray or high-resolution computed tomography. At baseline, median daily prednisone dose was 5 mg/day, and 13 patients (12.5%) received >15 mg/day of prednisone. The patients were mainly treated with methotrexate (38%) and mycophenolate mofetil (35.1%) at baseline, 9% of them receiving a combination of immunosuppressive therapy. Over follow-up, 42 patients (20.2%) switched to another immunosuppressive treatment, with no significant between-group difference.

Primary outcome analyses. Differences in mRSS at 12 ± 3 months were similar between the two groups (treated group: decrease of 2.7 [95% confidence interval {95% CI} 1.4–4.0] vs control group: 3.1 [95% CI 1.9–4.4], $P = 0.64$) (see Figure 2 and Table 2). Results were similar when considering patients with missing mRSS at 12 ± 3 months in the analysis (see Supplementary Table 2), when the analysis was restricted to the subgroup of patients with a shorter disease duration (≤24 months; see Supplementary Table 3), and when considering patients with baseline mRSS inferior to 22 (see Supplementary Table 4).

Secondary outcome analyses. Within the previous matched cohort of 208 patients, no significant difference was observed between control and treated groups for all secondary outcomes (see Table 2). When performing matched analysis for each of these outcomes, we found no effect of treatment on progressive skin fibrosis at 12 ± 3 months (odds ratio [OR] 1.3 [95% CI 0.6–3.0], $P = 0.53$) or on progressive lung fibrosis (OR 1.0 [95% CI 0.5–2.0], $P = 1$), progressive skin and lung fibrosis (OR 1.3 [95% CI 0.7–2.4], $P = 0.43$), or renal crisis (OR 1.0 [95% CI 0.1–7.2], $P = 1$; see Supplementary Table 5). In line with these findings, we did not find any difference in the subgroups of patients with missing outcomes at 12 ± 3 months (see Supplementary Table 5), in those with a shorter disease duration (≤24 months) (see Supplementary Table 6), or in those with a baseline mRSS <22 (see Supplementary Table 7).

DISCUSSION

This target trial emulation study showed no difference on skin fibrosis progression between patients with early dcSSc treated with immunosuppression monotherapy and those receiving immunosuppression plus oral glucocorticoids. In both groups, we observed a similar improvement in skin induration at one year with no clinically or statistically significant between-group difference. Two cases of scleroderma renal crisis occurred, one in each treatment group. The proportion of patients with progressive skin disease, progressive lung disease, and regressive skin and lung disease was similar in patients receiving or not receiving glucocorticoids.

Data from randomized trials and observational studies reveal widespread use of glucocorticoids in patients with dcSSc (40% of patients with dcSSc from the EUSTAR database were on

Table 1. Patient characteristics at baseline*

	Overall value	Control group		Treated group		*P*
		Value	% missing data	Value	% missing data	
N patients	208	104		104		
Male, n (%)	68 (33)	36 (35)	0.0	32 (31)	0.0	0.66
Age, y, mean ± SD	49 ± 12	50 ± 14	0.0	49 ± 11	0.0	0.97
Smoking ever, n (%)	62 (36)	31 (38)	22	31 (35)	14	0.24
Immunosuppressant treatment, n (%)						
Methotrexate	84 (40.4)	45 (43.3)	0.0	39 (37.5)	0.0	0.46
Rituximab	20 (9.6)	13 (12.5)	0.0	7 (6.7)	0.0	0.24
Cyclophosphamide	43 (20.7)	16 (15.4)	0.0	27 (26.0)	0.0	0.06
Mycophenolate mofetil	73 (35.1)	38 (36.5)	0.0	35 (33.7)	0.0	0.76
Other	6 (2.9)	4 (3.8)	0.0	2 (1.9)	0.0	0.62
Combination of immunosuppressant treatment, n (%)	19 (9.1)	12 (11.5)	0.0	7 (6.7)	0.0	0.30
Daily prednisone dose, mg/day, median (IQR)	–	–	–	5.0 (5.0–10.0)	5.8	
Disease duration, y, mean ± SD	2.4 ± 1.4	2.5 ± 1.3	0.0	2.3 ± 1.5	0.0	0.34
Baseline mRSS, median (IQR)	18.0 (12.0–23.0)	18.0 (12.0–23.0)	0.0	19.0 (12.8–23.0)	0.0	0.76
Forced vital capacity (% predicted), mean ± SD	84 ± 20	84 ± 22	14	84 ± 19	11	0.99
Diffusing capacity of the lung for carbon monoxide, mean ± SD	62 ± 18	64 ± 21	20	59 ± 15	18	0.11
Joint synovitis, n (%)	37 (18)	15 (15)	4	22 (22)	2	0.16
Anti-topoisomerase positive, n (%)	112 (59)	54 (56)	8	58 (61)	9	0.75
Anti-RNA polymerase III positive, n (%)	30 (25)	17 (30)	44	13 (20)	38	0.41
Presence of tendon friction rubs, n (%)	33 (17)	15 (16)	8	18 (18)	5	0.83
Interstitial lung disease[a]	100 (43.5)	51 (40.5)	0.0	49 (47.1)	0.0	0.38
Baseline year, n (%)			0.0		0.0	0.83
2013	36 (17)	20 (19)	–	16 (15)	–	–
2014	45 (22)	19 (18)	–	26 (25)	–	–
2015	44 (21)	20 (19)	–	24 (23)	–	–
2016	25 (12)	13 (12)	–	12 (11)	–	–
2017	17 (8)	10 (10)	–	7 (7)	–	–
2018	33 (16)	17 (16)	–	16 (15)	–	–
2019	8 (4)	5 (5)	–	3 (3)	–	–

* *P* values are provided by paired *t*-test for continuous variables, paired Wilcoxon tests for the mRSS, and McNemar test for categorical variables with two categories; the Friedman test was used for categorical variables with more than two categories. IQR, interquartile range; mRSS, modified Rodnan skin score.
[a] At lung x-ray or high-resolution computed tomography.

prednisone in 2018[15]). Although it is widely accepted that a short course of low-dose prednisone can help with fatigue, musculoskeletal pain, or itch,[5,6,9,13,33,34] its effectiveness for controlling skin fibrosis is more controversial. Even among experts, glucocorticoid prescribing practices for this indication vary widely.[8] Recent research intended to address this important point but could not provide a definite answer. The PRedSS study, a phase II randomized controlled trial of oral prednisolone in early diffuse cutaneous systemic sclerosis, converted to open label and terminated early during the COVID-19 pandemic,[13] assessed versus placebo the impact of adding a moderate dose of prednisolone (0.3 mg/Kg) to immunosuppression on skin fibrosis and disability at three months (two co-primary end points). The two co-primary end points were not met, but the study was underpowered because only 35 patients could be randomized. In this context of uncertainty, we designed this emulated trial to answer research questions that are difficult to study in randomized studies. Emulated target trials are particularly useful when the randomized study that would answer our causal question—the target trial—is not feasible, ethical, or timely.[16] We found no benefit of adding oral glucocorticoids to immunosuppression on skin fibrosis progression. We examined this aspect in a large sample of patients with a disease duration ≤5 years, as well as in those with a shorter disease duration (≤24 months) or with an mRSS between 7 and 22 who are potentially at higher risk to progress and more responsive to anti-inflammatory treatment.[21,25,26,28] In all cases, we did not observe any significant impact of prednisone on skin, suggesting that, if an effect of glucocorticoids exists, it is very likely below clinical significance.[35] Although our study has design limitations, it strongly questions the utility of prescribing oral glucocorticoids at the dose investigated for better skin fibrosis control in patients with early dcSSc.

Our data seem to suggest that the use of low-dose prednisone in early dcSSc is not a major risk factor for scleroderma renal crisis. In this study, in which patients started with a median prednisone dose of 5 mg/day, scleroderma renal crisis occurred in two patients, of whom only one was in the glucocorticoid-treated group. Our data are in keeping with previous literature[36] and with the results from Griffiths-Jones et al, in which no case of scleroderma renal crisis was observed in 17 patients with early dcSSc

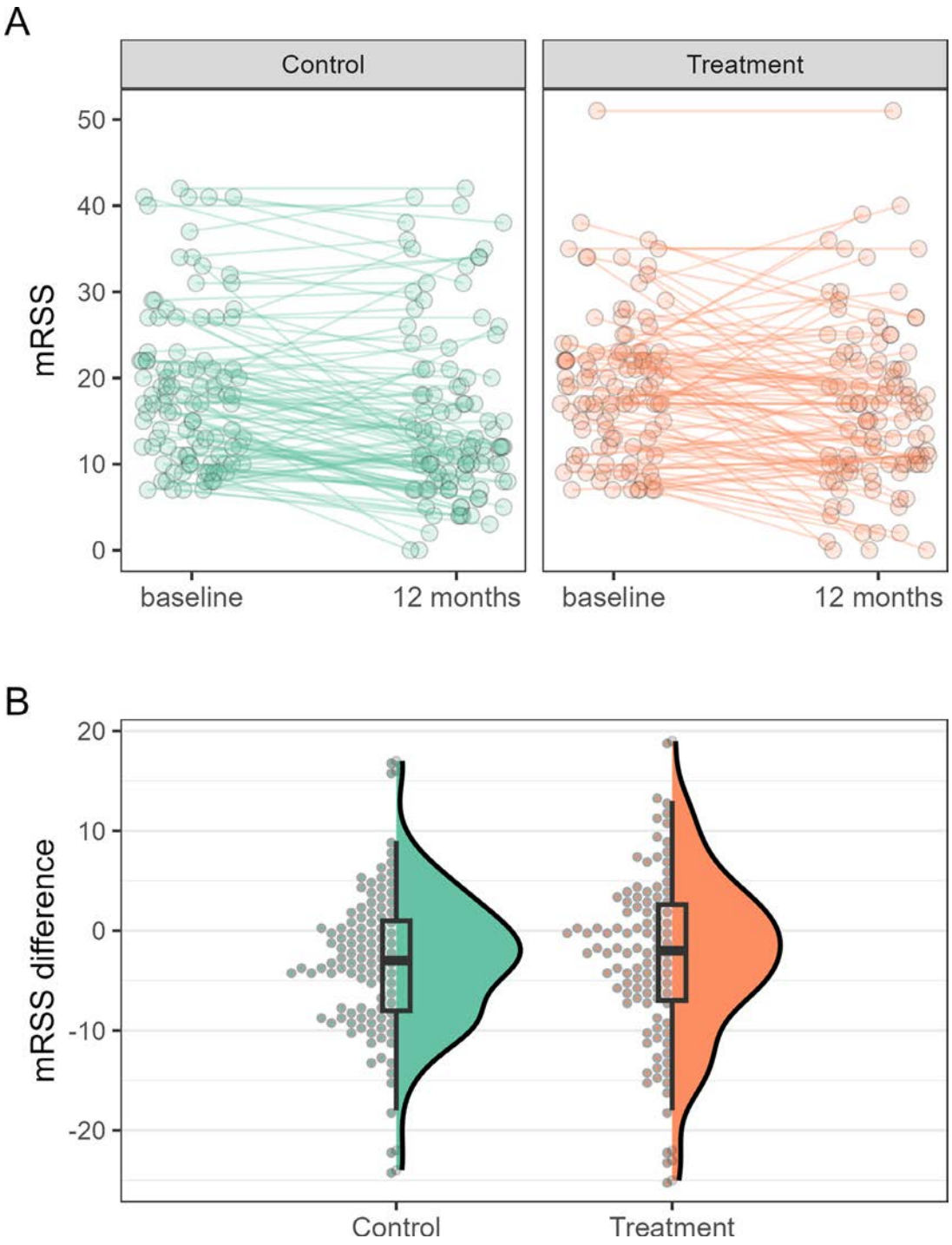


Figure 2. (A) Modified Rodnan skin score (mRSS) at baseline and after 12 ± 3 months for patients receiving immunosuppression alone ("control") and those receiving immunosuppression plus glucocorticoids ("treatment") is shown. (B) Distribution of the mRSS difference between 12 ± 3 months and baseline for the two groups is shown.

treated with daily 0.3 mg/Kg prednisolone.[13] However, our results should be interpreted with caution because our sample size may not have been sufficiently powered to answer this research question.

About one-third of patients from both treatment groups experienced progressive lung fibrosis. This suggests that adding prednisone to immunosuppression may not be beneficial for interstitial lung disease in early dcSSc. However, our study's primary focus was not on assessing lung outcomes, and adjustments for confounding were primarily tailored to the main primary end point. Therefore, these results should be interpreted with caution. Further research is required to address this clinically significant question.

Table 2. Primary and secondary outcomes for the main study*

		Controls		Treatment		
	Overall value	Value	% missing data	Value	% missing data	*P*
Primary outcome						
mRSS difference[a]	−2.9 (7.4)	−3.2 (6.8)	0	−2.6 (8.0)	0	0.55
Secondary outcomes						
Progressive skin fibrosis, n (%)	25 (12)	10 (10)	0	15 (14)	0	0.39
Progressive lung fibrosis, n (%)	41 (27)	23 (29)	23	18 (26)	34	0.83
Regressive skin and lung fibrosis, n (%)	83 (49)	41 (47)	15	42 (51)	21	0.48
Renal crisis, n (%)	2 (1)	1 (1)	0	1 (1)	0	0.99

* *P* values are provided by paired *t*-test for continuous variables and McNemar test for categorical variables. mRSS, modified Rodnan skin score.
[a] Between-group mean difference of mRSS at 12 ± 3 months from baseline.

Our study has major strengths. The EUSTAR database collects the clinical characteristics and outcomes of a large number of patients, enabling analysis of a rare disease phenotype difficult to target in a randomized controlled trial. Analyses benefited from multiple methods to address potential confounding by indication and subgroup, and sensitivity analyses were concordant for the primary outcome results, bolstering confidence in the main results. The international composition of the database allowed inclusion of patients from different countries potentially heterogeneous for treatment, racial background, and disease severity. This study summarizes the results of a potential pragmatic trial in which patients have received treatment in routine care setting. As a result, we were able to include patients likely to have more severe disease, multiple comorbidities, and possibly greater treatment resistance—patients who are generally less represented in randomized trials.

Some limitations need to be acknowledged. First, as in any observational analysis, assignment to a particular treatment was not randomized. Thus, residual confounding cannot be entirely ruled out. For example, although patients were matched for all known factors associated with skin fibrosis progression in early diffuse SSc, a higher proportion in the treatment group received cyclophosphamide. Although this could indicate greater disease severity, it may also reflect differences in physician experience, health care system, treatment availability, and cost considerations. However, it is important to note that the adjustments made in the analysis aimed to minimize such potential effect, and all subgroup analyses were consistent with the main finding.

Second, selection bias could play a role because we studied patients with information on skin involvement and disease duration. Third, we had no information about the precise duration, dose modification over the study period, and adherence to glucocorticoid intake in our sample. Although this imprecision may represent information bias, it also enhances the generalizability of our findings because our aim was to emulate a pragmatic trial in which no additional measures to assess treatment adherence are conducted, and the analysis was performed as an intention to treat. Intention-to-treat analysis helps minimize selection bias by including all participants as originally assigned, regardless of whether they strictly adhere to the treatment protocol. This approach accounts for participants who may switch between experimental and control groups or prematurely discontinue treatment, ensuring that the results reflect real-world variations in treatment adherence.[37] Moreover, the EUSTAR database does not provide information about the reasons leading physicians to prescribe treatment (either glucocorticoids or immunosuppression), limiting our ability to determine whether the outcome investigated (skin fibrosis) was intended to be targeted by the drugs prescribed. Another potential limitation is that we considered exploring the efficacy of glucocorticoid monotherapy but were unable to do so because of the very low number of patients receiving this treatment in the EUSTAR database. Finally, we were not able to assess the impact of glucocorticoid treatment on other patient-relevant outcomes such as pain, disability, or quality of life because of unavailable data.

In conclusion, the results of this target trial emulation study on patients with early dcSSc seen in routine care showed no benefit on skin of adding oral (≤20 mg/day) glucocorticoids to immunosuppression. Our findings question the utility of prescribing a low dose of these compounds for this indication, thus adding further knowledge to a debated subject.[8,10,33,34,38] Whether a short course of low-dose steroids could contribute to improve fatigue, pain, and itching was not the object of the present study and remains to be further investigated. However, the lack of an excess of renal complication in glucocorticoid-treated patients and clinical experience provide supports for these drugs to be used for symptom control while emphasizing the crucial need for monitoring early signs of kidney disfunction. Our results should be interpreted with caution because of the limitations of our observational study design.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr Iudici had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Mongin, Courvoisier, Iudici.

Acquisition of data. Matucci-Cerinic, Walker, Distler, Becvar, Siegert, Ananyeva, Smith, Alegre-Sancho, Yavuz, Limonta, Riemekasten, Rezus, Vonk, Truchetet, Del Galdo, Iudici.
Analysis and interpretation of data. Mongin, Courvoisier, Iudici.

REFERENCES

1. Meier FMP, Frommer KW, Dinser R, et al; EUSTAR Co-authors. Update on the profile of the EUSTAR cohort: an analysis of the EULAR Scleroderma Trials and Research group database. Ann Rheum Dis 2012;71(8):1355–1360.
2. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988; 15(2):202–205.
3. Herrick AL, Assassi S, Denton CP. Skin involvement in early diffuse cutaneous systemic sclerosis: an unmet clinical need. Nat Rev Rheumatol 2022;18(5):276–285.
4. van Laar JM, Farge D, Sont JK, et al; EBMT/EULAR Scleroderma Study Group. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial. JAMA 2014;311(24):2490–2498.
5. Denton CP, Hughes M, Gak N, et al; BSR and BHPR Standards, Guidelines and Audit Working Group. BSR and BHPR guideline for the treatment of systemic sclerosis. Rheumatology (Oxford) 2016; 55(10):1906–1910.
6. 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.
7. Galdo FD, Lescoat A, Conaghan PG, et al. OP0234 2023 update of EULAR recommendations for the treatment of systemic sclerosis. Ann Rheum Dis 2023;82(Suppl 1):154–155.
8. Fernández-Codina A, Walker KM, Pope JE; Scleroderma Algorithm Group. Treatment algorithms for systemic sclerosis according to experts. Arthritis Rheumatol 2018;70(11):1820–1828.
9. Hunzelmann N, Moinzadeh P, Genth E, et al. High frequency of corticosteroid and immunosuppressive therapy in patients with systemic sclerosis despite limited evidence for efficacy. Arthritis Res Ther 2009;11(2):R30.
10. Iudici M, van der Goes MC, Valentini G, et al. Glucocorticoids in systemic sclerosis: weighing the benefits and risks - a systematic review. Clin Exp Rheumatol 2013;31(2 Suppl 76):157–165.
11. Calguneri M, Apras S, Ozbalkan Z, et al. The efficacy of oral cyclophosphamide plus prednisolone in early diffuse systemic sclerosis. Clin Rheumatol 2003;22(4–5):289–294.
12. Sharada B, Kumar A, Kakker R, et al. Intravenous dexamethasone pulse therapy in diffuse systemic sclerosis. A randomized placebo-controlled study. Rheumatol Int 1994;14(3):91–94.
13. Griffiths-Jones DJ, Garcia YS, Ryder WD, et al. A phase II randomised controlled trial of oral prednisolone in early diffuse cutaneous systemic sclerosis (PRedSS). Rheumatology (Oxford) 2023;62(9):3133–3138.
14. Iudici M, Fasano S, Iacono D, et al. Prevalence and factors associated with glucocorticoids (GC) use in systemic sclerosis (SSc): a systematic review and meta-analysis of cohort studies and registries. Clin Rheumatol 2014;33(2):153–164.
15. Iudici M, Mongin D, Siegert E, et al; EUSTAR collaborators. Glucocorticoids prescribing practices in systemic sclerosis: an analysis of the EUSTAR database. Rheumatology (Oxford) 2023;62(4):1559–1567.
16. Dickerman BA, García-Albéniz X, Logan RW, et al. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med 2019;25(10):1601–1606.
17. Matthews AA, Danaei G, Islam N, et al. Target trial emulation: applying principles of randomised trials to observational studies. BMJ 2022; 378:e071108.
18. Walker UA, Tyndall A, Czirják L, et al. Clinical risk assessment of organ manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials And Research group database. Ann Rheum Dis 2007; 66(6):754–763.
19. 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.
20. Galluccio F, Walker UA, Nihtyanova S, et al. Registries in systemic sclerosis: a worldwide experience. Rheumatology (Oxford) 2011; 50(1):60–68.
21. Maurer B, Graf N, Michel BA, et al. Prediction of worsening of skin fibrosis in patients with diffuse cutaneous systemic sclerosis using the EUSTAR database. Ann Rheum Dis 2015;74(6):1124–1131.
22. Khanna D, Clements PJ, Furst DE, et al; Relaxin Investigators and the Scleroderma Clinical Trials Consortium. Recombinant human relaxin in the treatment of systemic sclerosis with diffuse cutaneous involvement: a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2009;60(4):1102–1111.
23. Pope JE, Bellamy N, Seibold JR, et al. A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. Arthritis Rheum 2001;44(6):1351–1358.
24. Hoffmann-Vold AM, Allanore Y, Alves M, et al; EUSTAR collaborators. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. Ann Rheum Dis 2021;80(2):219–227.
25. Wu W, Jordan S, Becker MO, et al. Prediction of progression of interstitial lung disease in patients with systemic sclerosis: the SPAR model. Ann Rheum Dis 2018;77(9):1326–1332.
26. Herrick AL, Peytrignet S, Lunt M, et al. Patterns and predictors of skin score change in early diffuse systemic sclerosis from the European Scleroderma Observational Study. Ann Rheum Dis 2018;77(4): 563–570.
27. Amjadi S, Maranian P, Furst DE, et al; Investigators of the D-Penicillamine, Human Recombinant Relaxin, and Oral Bovine Type I Collagen Clinical Trials. Course of the modified Rodnan skin thickness score in systemic sclerosis clinical trials: analysis of three large multicenter, double-blind, randomized controlled trials. Arthritis Rheum 2009;60(8):2490–2498.
28. Dobrota R, Maurer B, Graf N, et al; EUSTAR coauthors. Prediction of improvement in skin fibrosis in diffuse cutaneous systemic sclerosis: a EUSTAR analysis. Ann Rheum Dis 2016;75(10):1743–1748.
29. Mihai C, Dobrota R, Assassi S, et al. Enrichment strategy for systemic sclerosis clinical trials targeting skin fibrosis: a prospective, multiethnic cohort study. ACR Open Rheumatol 2020;2(8):496–502.
30. Rubin DB. Multiple Imputation for Nonresponse in Surveys. Wiley; 1987.
31. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2019. https://www.r-project.org
32. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45:1–67.
33. Blagojevic J, Legendre P, Matucci-Cerinic M, et al. Is there today a place for corticosteroids in the treatment of scleroderma? Autoimmun Rev. 2019;18(12):102403.
34. Iudici M. What should clinicians know about the use of glucocorticoids in systemic sclerosis? Mod Rheumatol 2017;27(6):919–923.
35. Khanna D, Furst DE, Hays RD, et al. Minimally important difference in diffuse systemic sclerosis: results from the D-penicillamine study. Ann Rheum Dis 2006;65(10):1325–1329.
36. Steen VD, Medsger TA Jr. Case-control study of corticosteroids and other drugs that either precipitate or protect from the development of scleroderma renal crisis. Arthritis Rheum 1998;41(9):1613–1619.

37. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017;18(6):1075–1078.
38. van den Hombergh WMT, Kersten BE, Knaapen-Hans HKA, et al. Hit hard and early: analysing the effects of high-dose methylprednisolone on nailfold capillary changes and biomarkers in very early systemic sclerosis: study protocol for a 12-week randomised controlled trial. Trials 2018;19(1):449.

APPENDIX A: EUSTAR COLLABORATORS

The EUSTAR collaborators are listed as follows: Florenzo Iannone (Bari, Italy); Otylia Kowal-Bielecka (Bialystok, Poland); Carmen Pizzorni (Genova, Italy); Francesco Ciccia (Naples, Italy); Simona Rednic (Cluj-Napoca, Romania); P. Vlachoyiannopoulos (Athens, Greece); Jiri Stork (Prague, Czech Republic); Murat Inanc (Capa, Istanbul, Turkey); Patricia E. Carreira (Madrid, Spain); Srdan Novak (Rijeka, Croatia); László Czirják (Pecs, Hungary); Eugene J. Kucharz (Katowice, Poland); Katja Perdan-Pirkmajer (Ljublijana, Slovenia); Bernard Coleiro (Balzan, Malta); Gianluca Moroncini (Ancona, Italy); Dominique Farge Bancel (Paris, France); Fabian A. Mendoza (Philadelphia, Pennsylvania, USA) Roger Hesselstrand (Lund, Sweden); Mislav Radic (Split, Croatia); Alexandra Balbir-Gurman (Haifa, Israel); Andrea Lo Monaco (Ferrara, Italy); Raffaele Pellerito (Torino, Italy); Alessandro Giollo (Verona, Italy); Jadranka Morovic-Vergles (Zagreb, Croatia); Christopher Denton (London, United Kingdom); Nemanja Damjanov (Belgrade, Serbia and Montenegro); Jörg Henes (Tübingen, Germany); Vera Ortiz Santamaria (Granollers, Barcelona, Spain); Stefan Heitmann (Stuttgart, Germany); Dorota Krasowska (Lublin, Poland); Paul Hasler (Aarau, Switzerland); Barbara Russo (Geneva, Switzerland); Michaela Kohm (Frankfurt am Main, Germany); Ivan Foeldvari (Hamburg, Germany); Gianluigi Bajocchi (Reggio Emilia, Italy); Maria Joao Salvador (Coimbra, Portugal); Bojana Stamenkovic (Niska Banja, Serbia and Montenegro); Carlo Francesco Selmi (Rozzano, Milano, Italy); Mohammed Tikly (Johannesburg, South Africa); Ariane Herrick (Salford, United Kingdom); Ulf Muller-Ladner (Bad Nauheim, Germany); Klaus Søndergaard (Aarhus N, Denmark); Francesco Puppo (Genova, Italy); Merete Engelhart (Hellerup, Denmark); Gabriela Szücs (Debrecen, Hungary); Carlos de la Puente (Madrid, Spain); Valeria Riccieri (Roma, Italy); Ruxandra Maria Ionescu (Bucharest, Romania); Ami Sha (Baltimore, Maryland, USA); Ana Maria Gheorghiu (Bucharest, Romania); Cord Sunderkötter (Münster, Germany); Jörg Distler (Erlangen, Germany); Francesca Ingegnoli (Milano, Italy); Luc Mouthon (Paris, France); Francesco Paolo Cantatore (Foggia, Italy); Susanne Ullman (Copenhagen, Denmark); Carlos Alberto von Muhlen (Porto Alegre, Brazil); Maria Rosa Pozzi (Monza, Italy); Kilian Eyerich (Munich, Germany); Piotr Wiland (Wroclaw, Poland); Marie Vanthuyne (Brussels, Belgium); Kristine Herrmann (Dresden, Germany); Ellen De Langhe (Leuven, Belgium); Branimir Anic, Marko Baresic, Miroslav Mayer (Zagreb, Croatia); Maria Üprus, Kati Otsa (Tallin, Estonia); Brigitte Granel (Marseille, France); Carolina de Souza Müller (Curitiba, Brazil); Svetlana Agachi (Chisinau, Republic of Moldova); Simon Stebbings (Dunedin, New Zealand); Alessandro Mathieu, Alessandra Vacca (Monserrato, CA, Italy); Percival D. Sampaio-Barros (São Paulo, Brazil); Lisa Stamp (Christchurch, New Zealand); Kamal Solanki (Hamilton, New Zealand); Douglas Veale (Dublin, Ireland); Esthela Loyo, Carmen Tineo (Santiago, Dominican Republic); Sergio Toloza (Catamarca, Argentina); Mengtao Li (Beijing, China); Walid Ahmed Abdel Atty Mohamed (Alexandria, Egypt); Jacek Olas (Crakow, Poland); Fahrettin Oksel, Figen Yargucu (Bornova, Izmir, Turkey); Cristina-Mihaela Tanaseanu (Bucharest, Romania); Rosario Foti (Catania, Italy); Codrina Ancuta (Iasi, Romania); Daniel E. Furst (Los Angeles, California, USA); Britta Maurer (Bern, Switzerland); Jacob van Laar (Middlesbrough, United Kingdom); Marzena Olesinska (Warsaw, Poland); Cristiane Kayser (São Paulo, Brazil); Nihal Fathi (Assiut, Egypt); Paloma Garcıa de la Pena Lefebvre, Jorge Juan Gonzalez Martin (Madrid, Spain); Patrick Carpentier, Bernard Imbert (Grenoble, France); Camille Francés, Patricia Senet (Paris, France); Jean Sibilia (Strasbourg, France); Ira Litinsky (Tel Aviv, Israel); Jean Luc Senécal, Martial Koenig, France Joval, Grodzicky Tamara (Montreal, Canada); Goda Seskute (Vilnius, Lithuania); Lesley Ann Saketkoo (New Orleans, Louisiana, USA); Eduardo Kerzberg (Buenos Aires, Argentina); Washington Bianchi, Breno Valdetaro Bianchi (Rio de Janeiro, Brazil); Ivan Castellvı (Barcelona, Spain); Jasminka Milas-Ahic, Roberta Visevic (Osijek, Croatia); Doron Rimar (Haifa, Israel); Maura Couto (Viseu, Portugal); François Spertini (Lausanne, Switzerland); Antonella Marcoccia (Roma, Italy); Sarah Kahl (Bad Bramstedt, Germany); Vivien M. Hsu (New Brunswick, New Jersey, USA); Thierry Martin (Strasbourg, France); Sergey Moiseev, Pavel Novikov (Moscow, Russia); Lorinda S. Chung (Stanford, Connecticut, USA); Tim Schmeiser (Wuppertal Elberfeld, Germany); Dominik Majewski (Poznan, Poland); Zbigniew Zdrojewski (Gdansk, Poland); Julia Martınez Barrio (Madrid, Spain); Dinesh Khanna (Ann Arbor, Michigan, USA); Vera Bernardino (Lisboa, Portugal); Lelita Santo (Coimbra, Portugal); Yair Levy (Kfar Saba, Israel); Omer Nuri Pamuk (Edirne, Turkey); Daniel Brito de Araujo (Pelotas, RS, Brazil); Piercarlo Sarzi Puttini (Milano, Italy); Marek Brzosko (Szczecin, Poland); Hadi Poormoghim (Tehran, Iran); Marta Maman (Buenos Aires, Argentina); Ina Kötter (Hamburg, Germany); Giovanna Cuomo (Naples, Italy); Francis Gaches (Toulouse, France); Laura Belloli (Milano, Italy); Petros Sfikakis (Athens, Greece); Juliana Markus (Uberlandia, Brazil); Daniel Furst (Los Angeles, California, USA); AnaMaria Ramazan (Constanta City, Romania); Patrick Jego (Rennes, France); Lorenzo Dagna (Milano, Italy); JM van Laar (Utrecht, the Netherlands); Lidia Rudnicka (Warsaw, Poland); Susana Oliveira (Amadora, Portugal); Fabiola Atzeni (Messina, Italy); Masataka Kuwana (Tokyo, Japan); Arsene Mekinian (Paris, France); Mickaël Martin (Poitiers, France); Yoshiya Tanaka (Kitakyushu, Japan); Hidekata Yasuoka (Aichi, Japan); Carmen-Pilar Simeon-Aznar (Barcelona, Spain); Tatsuya Atsumi (Sapporo, Japan); Magda Parvu (Bucharest, Romania); Ines Cordeiro (Lisboa, Portugal); Nicoletta Del Papa (Milano, Italy); Thomas Karonitsch (Vienna, Austria); Anna Bazela-Ostromecka (Grunwaldzka, Poland); Enrico Selvi (Siena, Italy); Yasushi Kawaguchi (Tokyo, Japan); Tomas Soukup (Hradec Kralove, Czech Republic); Ignasi Rodriguez-Pinto (Barcelona, Spain); Marija Geroldinger-Simic (Linz, Austria); Gerard Espinosa (Barcelona, Spain); Karen Voigt (Hamburg, Germany); Torsten Kubacki (Köln, Germany); Olena Garmish (Kiev, Ukraina); Marta Mosca (Pisa, Italy); Ulrich Gerth (Rheinfelden, Switzerland); Ludmila Antonenko (Kiev, Ukraina).

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 658–665
DOI 10.1002/acr.25470

Musculoskeletal Ultrasound Practices of Graduates of a Blended-Learning Program: A Survey of Rheumatologists From the United States

Midori Nishio,[1] Karina D. Torralba,[2] Sonja I. Ziniel,[3] Eugene Kissin,[4] and Fawad Aslam,[5]
for the Ultrasound School of North American Rheumatologists

Objective. Not much is known regarding musculoskeletal ultrasound (MSUS) practices of rheumatologists in the United States. We sought to determine the current use of MSUS among past participants of the Ultrasound School of North American Rheumatologists (USSONAR) training program and, by extension, MSUS-practicing rheumatologists and to understand barriers to its MSUS use.

Methods. An online survey was sent to 374 participants in the eight-month USSONAR blended course (Fundamentals in Musculoskeletal Ultrasound and Train the Trainer) between 2009 and 2020. Each respondent had a unique identifier linked to their total number of submitted practice scans and examination scores during training.

Results. The survey response rate was 28.1% (105 of 374), comprising 82% adult and 18% pediatric rheumatologists. Of the respondents, 71% were MSUS certified: 86.7% performed and/or interpreted diagnostic MSUS, 81.0% performed/interpreted procedural MSUS, 59.8% billed for at least 50% of diagnostic studies, and 78.8% billed for at least 50% of procedural studies. The top reasons for not doing diagnostic and procedural ultrasonography were lack of administrative support and limited time, respectively. For 25% of diagnostic ultrasonography and 12.9% of procedural ultrasonography, billing was done <50% of the time. Of the respondents, 78.0% reported that USSONAR training made them better rheumatologists.

Conclusion. Most USSONAR-trained rheumatologists are certified, practicing both diagnostic and procedural MSUS and billing for most of their work. However, a substantial number of studies are not being billed due to time constraints, limited administrative support, and legal liability. Participants agreed that USSONAR training made them better rheumatologists.

INTRODUCTION

Musculoskeletal ultrasound (MSUS) was first used in 1972 to differentiate a Baker cyst from thrombophlebitis.[1] The earliest rheumatology-specific application was reported in 1978, when MSUS was used to detect synovitis.[2] Since then, the use of MSUS has increased dramatically. Between 2000 and 2009, MSUS performed by rheumatologists increased more than 12,500%.[3] The widespread appeal and use of MSUS are evident because more nonradiologists now perform MSUS than radiologists.[4,5] The advantages of MSUS are its lack of radiation, lower cost, dynamic elements, and increasingly versatile and affordable equipment. However, MSUS remains an operator-dependent modality, and competence and proficiency are critical to obtaining accurate results. Competence is defined as an operator having the necessary combination of knowledge and skill to perform a procedure correctly.[6] Training to acquire the knowledge and skill for certification comes through different pathways.

Currently, in the United States, two certifications to demonstrate competence are available to rheumatology health professionals. One is the Musculoskeletal Ultrasound Certification in Rheumatology (RhMSUS) from the American College of

Presented at the American College of Rheumatology Convergence 2023, November 10 to 15, 2023, San Diego, California. Portions of this manuscript have been published in abstract form: *Arthritis Rheumatol* 2023;75(Suppl 9).

[1]Midori Nishio, MD: John Muir Specialty Medical Group, Walnut Creek, California; [2]Karina D. Torralba, MD: Loma Linda University School of Medicine, Loma Linda, California; [3]Sonja I. Ziniel, MA, PhD: University of Colorado School of Medicine and Children's Hospital Colorado, Aurora; [4]Eugene Kissin, MD: Boston University Medical Center, Boston, Massachusetts; [5]Fawad Aslam, MBBS, MS: Mayo Clinic, Scottsdale, Arizona.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25470).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25470.

Address correspondence via email to Fawad Aslam, MBBS, MS, at aslam.fawad@mayo.edu.

Submitted for publication March 29, 2024; accepted in revised form October 29, 2024.

SIGNIFICANCE & INNOVATIONS

- Not much is known about musculoskeletal ultrasound practices of adult and pediatric rheumatologists in the United States. This survey study attempts to address this gap.
- More than 80% of online survey respondents performed diagnostic and procedural musculoskeletal ultrasound.
- Although most billed for their work, a substantial number of musculoskeletal ultrasound examinations were not billed owing to lack of time, fear of legal liability, or insufficient reimbursement.
- Most practitioners felt that the training provided by the Ultrasound School of North American Rheumatologists made them a better rheumatologist.

Rheumatology (ACR); the other is the Registered in Musculoskeletal (RMSK) certification from the Alliance for Physician Certification and Advancement. In 2017, 94% of adult rheumatology fellowship programs in the United States included some form of MSUS training in their curricula.[7] Outside of fellowship, other training opportunities include online modules and MSUS conferences. Since 2008, the Ultrasound School of North American Rheumatologists (USSONAR), a not-for-profit organization, has participated in MSUS training for fellows and postgraduate rheumatologists in both adult and pediatric rheumatology.[6] Between 2009 and 2020, over 450 students graduated from the school's eight-month combination training program, Fundamentals in Musculoskeletal Ultrasound (FMU) and Train the Trainer (TTT).[8] The FMU participants were mainly fellows in rheumatology training programs, whereas the TTT participants could not be fellows in training and were mainly academic professionals in rheumatology. The latter course was supported by the ACR in its initial stages, and the program overall was the focus of a 2011 Rheumatology Research Foundation Clinician Scholar Educator Award given to Eugene Kissin, MD, the founder of the program. Not much is known about the perceptions and current practices of trained rheumatology health professionals using MSUS in the United States or about the USSONAR courses. Therefore, we conducted a study of previous USSONAR course participants to determine the current patterns of MSUS use among those participants and, by extension, MSUS-practicing rheumatologists as well as to understand barriers to MSUS use.

MATERIALS AND METHODS

Institutional review board approval for the study was obtained through Loma Linda University Medical Center (5220266). A statement giving consent was required to participate in the survey. This cross-sectional survey included 482 possible participants of the USSONAR FMU/TTT programs who completed a final examination between 2009 and 2020. An updated email list was created for the 482 participants by using email addresses provided to the USSONAR and by using online sources. In September 2022, a 28-question, 10-minute online survey (Qualtrics) was sent to the previous course participants. Two weekly reminders were sent two weeks apart, with survey closure eight weeks after the initial email invitation. The USSONAR leadership also sent reminders to personal email addresses (ie, email addresses not identified through the study search strategy) and personal telephone contact numbers, when known, to improve the survey response rate. There was no financial incentive to complete the survey. Respondents were given a unique identifier matched to their examination scores and the total number of practice studies (scans) submitted in their training year.

Data were collected for demographic and practice characteristics of the respondents. Survey questions focused on whether respondents were practicing diagnostic and/or procedural MSUS and if they were billing for these practices. We inquired about the obstacles limiting their MSUS practices. Some sonographic clinical scenarios, developed by subject matter experts, were also given to assess respondents' confidence in addressing them. The complete survey is included as Supplement 1. The survey was developed by experts in MSUS, and because of a lack of similar survey tools, the survey was based on the experts' experience. The questionnaire was developed with the help of a statistician who had expertise in survey design. The Checklist for Reporting Survey Studies tool for designing cross-sectional surveys was used in writing the manuscript and is provided as Supplement 2.[9] Two rounds of pretesting were done, and with the help of the survey statistician, subsequent revisions were made to avoid cognitive bias. Those participating in the pretests were excluded from the final survey.

Statistical analyses. Before analyzing survey data, we statistically assessed potential nonresponse bias by first determining if examination scores, the number of practice studies, or the program type were significant predictors of survey response and our main survey outcome of interest. We used a multiple logistic regression model to predict the likelihood of each participant responding to the survey. Relationships between the use of MSUS and examination scores as well as the number of practice studies were assessed through nonparametric median tests. χ^2 tests were used to determine if a relationship between program type and the use of MSUS existed. $P < 0.05$ was deemed significant. Survey data were only analyzed using descriptive statistics such as counts and percentages for categorical variables and medians and interquartile ranges for continuous variables. All analyses were performed using Stata/IC 16.1 (StataCorp [2020], Stata Statistical Software Release 16, StataCorp, LLC).

RESULTS

Of the 482 possible participants who were sent the survey, 108 emails could not be delivered, leaving 374 FMU and TTT

Table 1. Demographic and practice characteristics of respondents (n = 105)*

Characteristic	n (%)
Sex	
Female	57 (55.3)
Male	46 (44.7)
Age, median (IQR), y	41 (36–38)
Time practicing MSUS, median (IQR), y	6 (3.5–9)
Time practicing clinical rheumatology, median (IQR), %	90 (75–100)
Certification in MSUS or rheumatic ultrasonography (multiple selections are possible)	
RhMSUS	61 (61.0)
RMSK	6 (6.0)
Other	4 (4.0)
Never certified	31 (31.0)
Practice type	
Department of academic medical center, teaching institution	73 (70.9)
Department of nonacademic medical center, nonteaching	8 (7.8)
Multispecialty private practice, nonteaching	10 (9.7)
Large, single-specialty practice, nonteaching (more than three providers)	4 (3.9)
Small, single-specialty practice, nonteaching (three or fewer providers)	5 (4.9)
Other	3 (2.9)
Primary practice location in the United States	
Northeast	30 (29.4)
Midwest	19 (18.6)
South	27 (26.5)
West	25 (24.5)
Not in the United States	1 (1.0)
Patient population	
Primarily pediatric patients	19 (18.1)
Primarily adult patients	86 (81.9)

* There were the following numbers of respondents for the listed characteristics: sex (n = 103), age (n = 102), time practicing MSUS (n = 100), time practicing clinical rheumatology (n = 103), certification in MSUS or rheumatic ultrasonography (n = 100), across practice types (n = 103), across primary practice locations in the United States (n = 102), and all patient populations (n = 105). IQR, interquartile range; MSUS, musculoskeletal ultrasound; RhMSUS, Musculoskeletal Ultrasound Certification in Rheumatology; RMSK, Registered in Musculoskeletal.

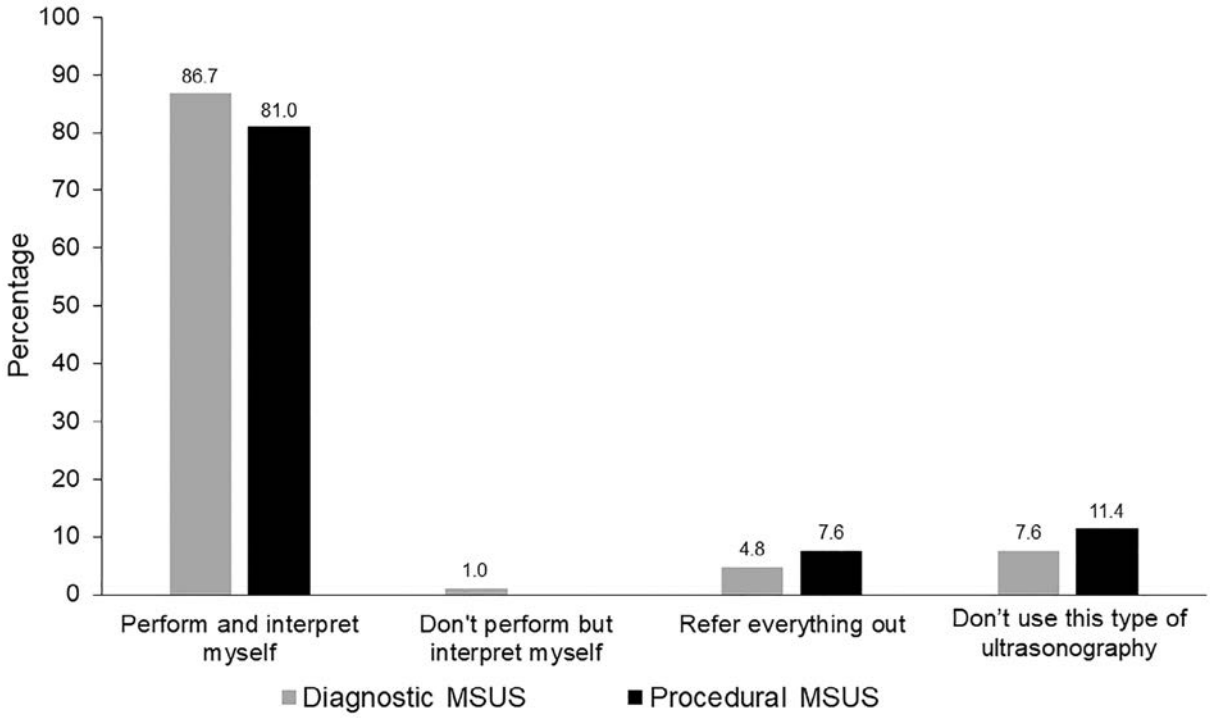


Figure 1. Percentage of respondents performing and/or interpreting diagnostic and/or procedural MSUS (n = 105). MSUS, musculoskeletal ultrasound.

Table 2. Reasons for not performing diagnostic or procedural MSUS*

Reasons[a]	Diagnostic MSUS, n (%)[b]	Procedural MSUS, n (%)[c]
Lack of confidence in my MSUS skills	3 (21.4)	10 (50.0)
Concerns about legal liability	3 (21.4)	3 (15.0)
Lack of administrative support	4 (28.6)	8 (40.0)
Opposition from other imaging groups (eg, radiology)	1 (7.1)	0 (0)
Lack of interest in diagnostic/procedural MSUS	0 (0)	0 (0)
Lack of time	3 (21.4)	14 (70.0)
Inadequate reimbursement	1 (7.1)	5 (25.0)
Limited benefit of diagnostic MSUS in clinical practice	0 (0)	NA
Lack of certification	2 (14.3)	2 (10.0)
Other	2 (14.3)	3 (15.0)

* MSUS, musculoskeletal ultrasound; NA, not applicable.
[a] Multiple selections were possible.
[b] For adult rheumatology, n = 14. For pediatric rheumatology, n = 1 (7.1%).
[c] For adult rheumatology, n = 20. For pediatric rheumatology, n = 6 (30.1%).

participants who received the survey. The overall response rate was 28.1% of participants (105 of 374). The course taken influenced survey participation: 43.5% of participants (40 of 92) for the TTT course versus 23.0% of participants (65 of 282) for the FMU course ($P < 0.005$). Analyses using background variables such as FMU versus TTT status, year of graduation, number of studies submitted, or the USSONAR course final examination scores did not show evidence of nonresponse bias (data not shown). Demographic characteristics are summarized in Table 1.

Respondents had practiced MSUS for a median (interquartile range) of 6 (3.5–9) years. Of the respondents, 70.9% practiced in academic medical centers, 71% held an MSUS certification, 81.9% were adult rheumatologists, and 18.1% were pediatric rheumatologists. In addition, 86.7% performed and/or interpreted diagnostic MSUS themselves, and 81.0% performed/interpreted their own procedures (Figure 1). No statistically significant difference was shown in the billing practices for those in academic versus nonacademic settings. Of all respondents combined, 59.8% billed for at least 50% of diagnostic studies, whereas 78.8% billed for at least 50% of their procedural studies.

The leading reason for not performing diagnostic MSUS was lack of administrative support. The leading reason for not performing procedural MSUS was lack of time (Table 2). Most respondents billed for their MSUS work. Only 15.2% of practitioners doing diagnostic work and 8.2% of those doing procedural work never billed for their services (Figure 2). The leading reason for not billing for MSUS work and/or for billing less than 50% of the time in up to 44.4% of responses was no time for documentation (44.4% in procedural and 43.2% in diagnostic MSUS), followed by poor reimbursement for effort (38.9% in procedural and 35.1% in diagnostic MSUS; Supplementary Table 1).

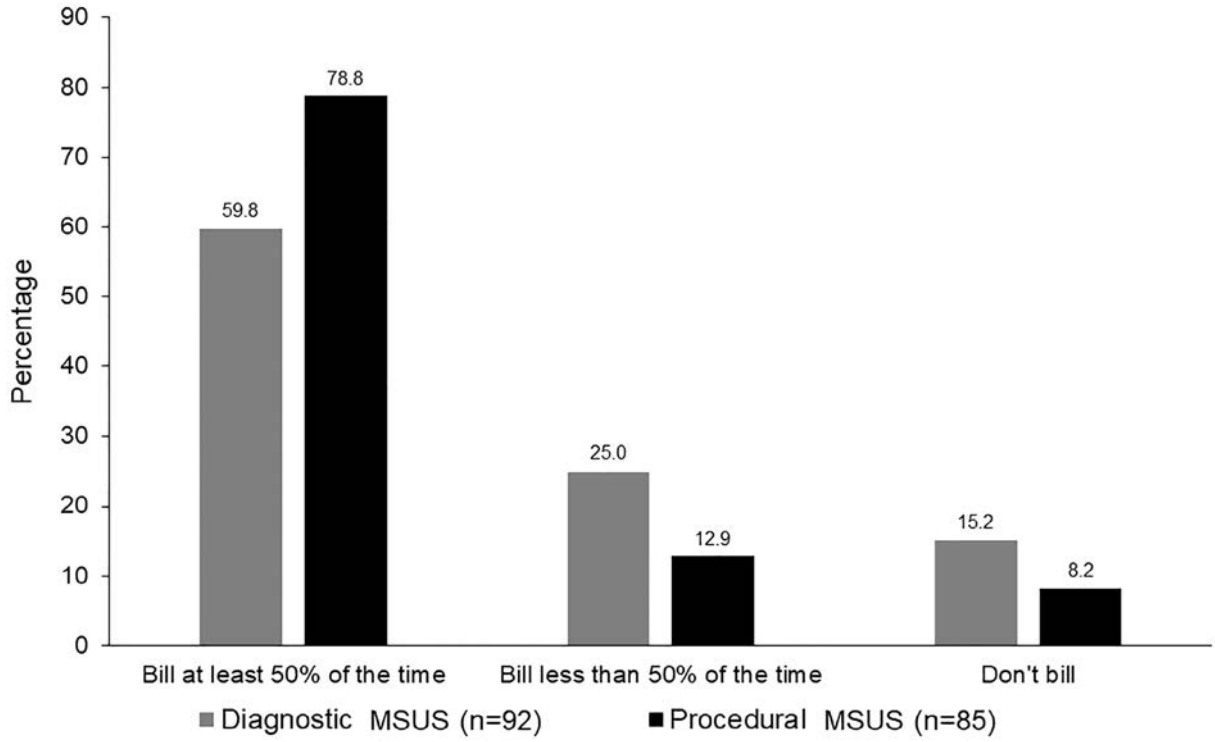


Figure 2. Billing for diagnostic and procedural MSUS. MSUS, musculoskeletal ultrasound.

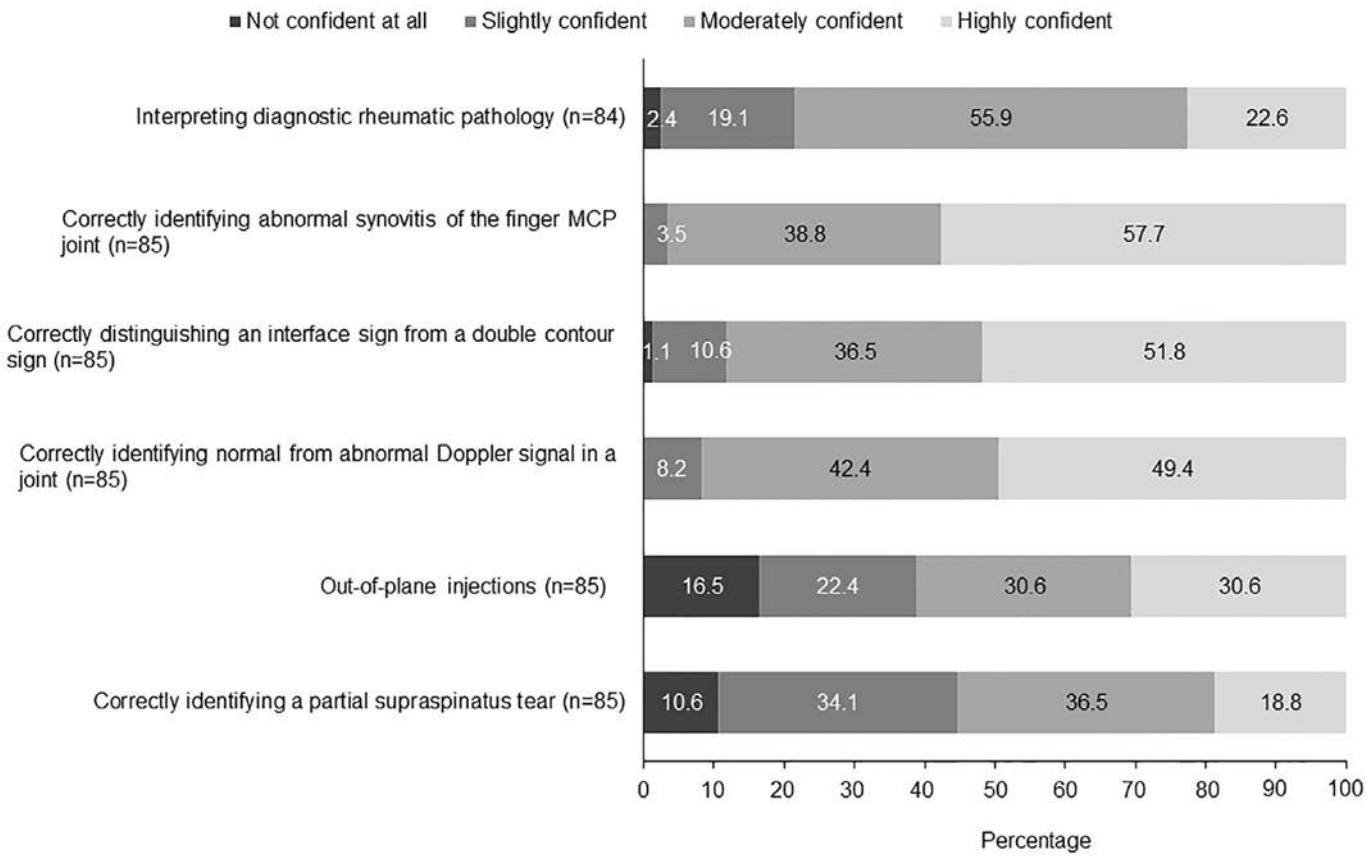


Figure 3. Confidence of adult practitioners in scenarios two years after participating in the Ultrasound School of North American Rheumatologists course. MCP, metacarpophalangeal.

When asked to rate their confidence in a set of MSUS skills two years after their USSONAR course, most of the adult rheumatologists (Figure 3) said they were moderately to highly confident in their skills. The out-of-plane injection scenario had the highest number of respondents (16.5%) with the lowest confidence. Similarly, most pediatric rheumatologists felt moderately confident in a set of pediatric MSUS skills (Figure 4). Of the given scenarios, adult clinicians were least confident about performing out-of-plane injections and identifying partial supraspinatus tendon tears, whereas pediatric clinicians were least confident about performing intra-articular hip injections. All responses were compared between the FMU and TTT groups, with no significant differences except for the TTT group being less confident in intra-articular hip injections ($P < 0.05$, Fisher exact test). About 78% of participants (80 of 103) reported that the USSONAR training had made them a better rheumatology practitioner.

DISCUSSION

This US-based survey (101 of 102 of the respondents practiced in the US) of USSONAR participants provides insights into current MSUS practices of this group of rheumatologists. The survey showed that most respondents were actively using MSUS. Nationally, use of MSUS, primarily to guide interventions, increased substantially across specialties.[4] It was encouraging to learn that both diagnostic (86.7%) and procedural (81.0%) MSUS were widely used by our study participants. In a similar survey of Portuguese rheumatologists (36% response rate), the use of diagnostic and procedural MSUS was 59% and 66%, respectively, which is lower than the rates in our study.[10] The two surveys are difficult to compare, however, because of the inherent differences in the two health systems. In addition, a 2016 survey of rheumatologists in Central and Eastern European countries supported less use of MSUS, especially for procedural guidance: only 11.6% of examinations in their survey had been done with MSUS for procedural guidance versus almost 60% for diagnostic purposes.[11]

Diagnostic MSUS is becoming increasingly important in rheumatology and is now part of classification criteria for gout, polymyalgia rheumatica, and calcium pyrophosphate deposition disease and has been proposed for assessing lupus inflammatory arthritis as part of a disease activity index.[12–15] From a patient- and practitioner-reported outcome perspective, recent data from the United States show that a dedicated rheumatology MSUS clinic facilitates medical decision-making for patients by improving

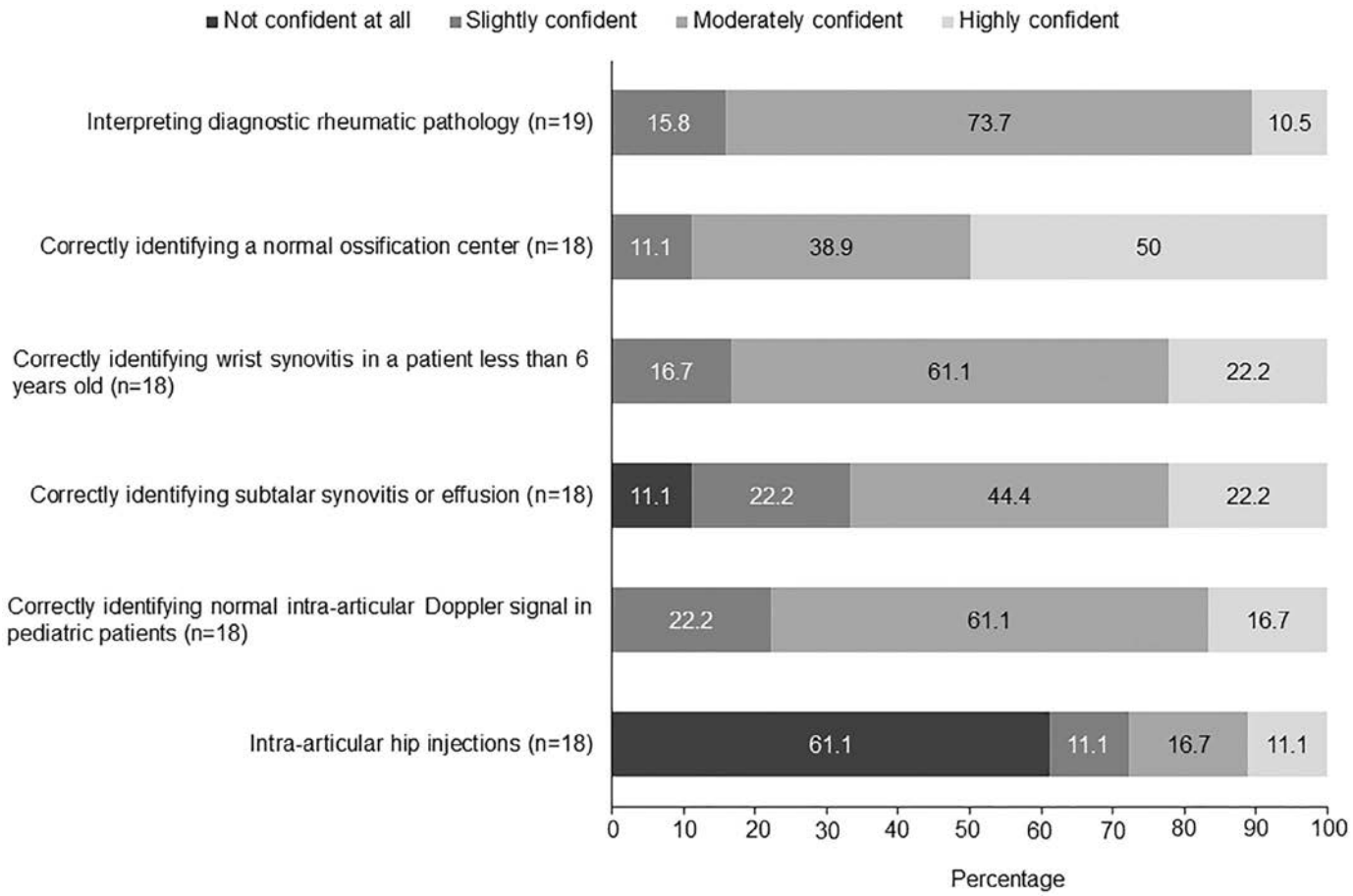


Figure 4. Confidence of pediatric practitioners in scenarios two years after participating in the Ultrasound School of North American Rheumatologists course.

understanding of their disease and by meaningfully influencing treatment decisions of referring practitioners.[16] MSUS is an important aid to clinical decision-making and to educating patients about a disease process.[11] Therefore, diagnostic MSUS should be encouraged, and efforts should be made to identify and address barriers to its successful adoption, which could save health care costs, as was originally envisioned.[3]

Most respondents were billing for their work at least 50% of the time. We were not surprised that the billing proportion for diagnostic MSUS was lower than that of procedural MSUS because the former requires more in-depth knowledge of MSUS, is time intensive, and has potentially greater legal liability. Of note, however, about 40% of respondents were billing <50% of the time for diagnostic examinations, and 20% were billing <50% of the time for procedural examinations, thereby forgoing revenue. This finding is concerning considering that 71% of the respondents were certified. Billing by rheumatologists for diagnostic MSUS is also complicated by current procedural terminology codes, which were determined by radiologists. The current definitions for complete and incomplete MSUS examinations by radiologists are based on anatomic regions, whereas the definition of a complete examination by a rheumatologist may involve the evaluation of selective joints from several anatomic areas to confirm the diagnosis of a polyarticular inflammatory disease and/or to evaluate disease activity.[17]

Revenue from MSUS can be an important contributor to a rheumatology department's net income.[18] The top reported reasons for not billing included limited time and lack of administrative support. This is an area for further research. For this study, "lack of administrative support" referred to the process of buying/leasing an ultrasonography machine, maintaining it, setting up compliant data archives, creating and proctoring a billing pathway, creating a patient schedule to allow for use of ultrasonography, marketing the availability of ultrasonography for rheumatology, and tracking the success of adding ultrasonography to a practice. However, the survey question only gave "lack of administrative support" as an option. No examples were given, leaving the question open to respondents' interpretation. These types of administrative activities are new to traditional clinic administrators, who need to be educated about the value of adding ultrasonography to a practice. For compliance, we had to keep the survey short, which prevented us from investigating this

topic further. In contrast to our study, the Portuguese rheumatology MSUS survey reported that access to equipment was the main limitation.[10] In that study, billing was not specifically queried, but 89% of respondents were writing complete MSUS reports. Again, direct comparisons were difficult due to health system differences. However, rheumatologists were reimbursed for performing MSUS in only 62% of member countries (21 of 34) in the EULAR.[19] When comparing our results with survey results from other countries, it is also important to note that their surveys often included participants from various courses, unlike the USSONAR. This difference may affect the responses.

For the clinical scenarios in the survey, adult and pediatric practitioners reported their lowest confidence for procedural scenarios, although the absolute percentages were small. One explanation could be that the Accreditation Council of Graduate Medical Education outlined the performance of ultrasonography by rheumatology fellows only in the second version of the Rheumatology Milestones (adult programs, August 2020; pediatric programs, April 2023) as part of patient care subcompetency for procedures (milestone PC4) but only at the level of the expert learner (level 5).[20,21] In contrast, milestones outlined for the specialties of physical medicine and rehabilitation[22] and sports medicine[23] specify performing MSUS at the earlier learner level of advanced beginner (level 2). The FMU and TTT programs through the USSONAR represent a finite approach to improving procedural skills, but if fellowship programs are not encouraging the adoption of these skills for early-level learners, then some deficiencies are likely to remain.

Most participants (78%) agreed with the statement that their USSONAR training made them better rheumatology providers. This suggests that the FMU/TTT USSONAR course curriculum provided adequate training and gave the participants confidence, as reflected by the large number of respondents with an MSUS certification (70%). However, of all MSUS-certified rheumatology providers (RhMSUS or RMSK), the percentage who are also USSONAR trained is unknown. The reasons for respondents not pursuing certification remain an important unknown and a topic for future research. The USSONAR provides a unique blended program consisting of online learning with volunteer faculty providing feedback on the participants' submitted MSUS scan examinations, web-based resources, in-person workshops, and a final composite assessment consisting of a written examination and an objective, structured clinical examination. The strength and success of the program rely on the participation of self-directed learners. For dedicated learners, competency has been shown to be achievable even without mentors providing in-person supervision.[24] The USSONAR program, therefore, successfully fills a major need of rheumatology MSUS training requirements. In Europe, the Spanish Society of Rheumatology has a long-standing program that offers an MSUS training program to its rheumatology trainees. A survey of rheumatology residents spanning 10 years (31% response rate) showed that 88% found the training program to be beneficial.[25] However, the attainment of competency accreditation/certification was much lower at 5%, in contrast to our survey's 70%. This comparison must be interpreted with the caveat of different training environments, systems, and regulations.

Our study has limitations. As is inherent in survey research, our low response rate was not unusual for a web-based survey[5,10,25,26]; however, nonresponse has also been reported as less of an issue in physician survey research.[27] Multiple attempts were made to contact previous participants in the FMU and TTT courses. We cannot predict how the nonrespondents or those without correct email addresses would have responded; 108 of 482 initial emails were undeliverable. Those who responded may have had a greater interest in MSUS than those who did not respond.[27] The nonrespondents may not have been actively practicing MSUS and, therefore, were not inclined to participate.[28] In addition, nonparticipation could have resulted from nonuse of email, concerns about privacy, incorrect contact information, and/or survey fatigue. Despite the low response rate, our internet searches for names and practice locations showed most participants in the courses to be practicing rheumatology providers. The response rate between the TTT and FMU track was significantly different statistically, which may have affected the responses and practice patterns. The results reflect practice patterns of USSONAR graduates only and not all rheumatology ultrasonography practitioners. Rather than relying on survey methodology, an alternative method for accurately establishing how many USSONAR graduates actively use ultrasound would be to analyze Medicare claims data for ultrasonography codes submitted by rheumatologists and to screen these against the USSONAR database.

From this survey study, we reported current MSUS practices of USSONAR graduate rheumatologists in the United States. Although not all US-based rheumatologists performing ultrasonography are USSONAR graduates, the USSONAR is the leading MSUS training organization for rheumatologists in the United States. Most respondents were certified, practicing both diagnostic and procedural MSUS and billing for their examinations at least 50% of the time. However, a substantial number of MSUS studies are still not being billed for lack of time, fear of legal liability, and/or a feeling of inadequate return on start-up costs for MSUS. Most participants agreed that USSONAR training made them better rheumatologists.

ACKNOWLEDGMENTS

The authors thank Marianne Mallia, ELS, MWC, senior scientific/medical editor, Mayo Clinic, for substantive edits to the manuscript. They also thank the Scientific Publications staff at the Mayo Clinic for providing proofreading, administrative, and clerical support.

AUTHOR CONTRIBUTIONS

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

version to be published. Dr Nishio had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Nishio, Torralba, Ziniel, Kissin, Aslam.

Acquisition of data. Nishio, Torralba, Kissin, Aslam.

Analysis and interpretation of data. Nishio, Torralba, Ziniel, Aslam.

REFERENCES

1. McDonald DG, Leopold GR. Ultrasound B-scanning in the differentiation of Baker's cyst and thrombophlebitis. Br J Radiol 1972;45(538):729–732.
2. Cooperberg PL, Tsang I, Truelove L, et al. Gray scale ultrasound in the evaluation of rheumatoid arthritis of the knee. Radiology 1978;126(3):759–763.
3. Sharpe RE, Nazarian LN, Parker L, et al. Dramatically increased musculoskeletal ultrasound utilization from 2000 to 2009, especially by podiatrists in private offices. J Am Coll Radiol 2012;9(2):141–146.
4. Sharpe RE Jr, Nazarian LN, Levin DC, et al. The increasing role of nonradiologists in performing ultrasound-guided invasive procedures. J Am Coll Radiol 2013;10(11):859–863.
5. Berko NS, Goldberg-Stein S, Thornhill BA, et al. Survey of current trends in postgraduate musculoskeletal ultrasound education in the United States. Skeletal Radiol 2016;45(4):475–482.
6. Kissin EY, Niu J, Balint P, et al. Musculoskeletal ultrasound training and competency assessment program for rheumatology fellows. J Ultrasound Med 2013;32(10):1735–1743.
7. Torralba KD, Cannella AC, Kissin EY, et al. Musculoskeletal ultrasound instruction in adult rheumatology fellowship programs. Arthritis Care Res (Hoboken) 2020;72(6):859–870.
8. USSONAR Training Program. Ultrasound School of North American Rheumatologists. Accessed January 26, 2024. https://www.ussonar.org/training-program
9. Sharma A, Minh Duc NT, Luu Lam Thang T, et al. A consensus-based Checklist for Reporting of Survey Studies (CROSS). J Gen Intern Med 2021;36(10):3179–3187.
10. Neto A, Figueiredo G, Saraiva F, et al. Musculoskeletal ultrasound among rheumatologists in Portugal: state of practice and training. Acta Reumatol Port 2020;45(2):152–154.
11. Mandl P, Baranauskaite A, Damjanov N, et al. Musculoskeletal ultrasonography in routine rheumatology practice: data from Central and Eastern European countries. Rheumatol Int 2016;36(6):845–854.
12. Abhishek A, Tedeschi SK, Pascart T, et al. The 2023 ACR/EULAR classification criteria for calcium pyrophosphate deposition disease. Arthritis Rheumatol 2023;75(10):1703–1713.
13. Dasgupta B, Cimmino MA, Kremers HM, et al. 2012 Provisional classification criteria for polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. Arthritis Rheum 2012;64(4):943–954.
14. Neogi T, Jansen TL, Dalbeth N, et al. 2015 Gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2015;74(10):1789–1798.
15. Sandler RD, Vital EM, Mahmoud K, et al. Revision to the musculoskeletal domain of the BILAG-2004 index to incorporate ultrasound findings. Rheumatology (Oxford) 2024;63:498–505.
16. Fairchild RM, Deluna MD, Golovko V, et al. Evolution and impact of a dedicated ultrasound clinic on clinical rheumatology practice at an academic medical center. Semin Arthritis Rheum 2023;63:152276.
17. Torralba KD, Choi KS, Salto LM, et al. Musculoskeletal ultrasound scanning protocol consensus statements on scanning conventions and documentation in the US. Arthritis Care Res (Hoboken) 2020;72(8):1177–1184.
18. D'Anna KM, Lynch CS, Cabling M, et al. Clinical academic rheumatology: a boon for health systems. Arthritis Care Res (Hoboken) 2022;74(7):1041–1048.
19. Mandl P, Ciechomska A, Terslev L, et al. Implementation and role of modern musculoskeletal imaging in rheumatological practice in member countries of EULAR. RMD Open 2019;5(2):e000950.
20. Accreditation Council for Graduate Medical Education. Pediatric Rheumatology Milestones. Accreditation Council for Graduate Medical Education; 2023. Accessed January 26, 2024. https://prod.acgme.org/globalassets/pdfs/milestones/pediatricrheumatologymilestones2.0.pdf
21. Accreditation Council for Graduate Medical Education. Rheumatology Milestones. Accreditation Council for Graduate Medical Education; 2020. Accessed January 26, 2024. https://www.acgme.org/globalassets/pdfs/milestones/rheumatologymilestones2.0.pdf
22. Accreditation Council for Graduate Medical Education. Physical Medicine and Rehabilitation Milestones. Accreditation Council for Graduate Medical Education; 2021. Accessed January 26, 2024. https://www.acgme.org/globalassets/pdfs/milestones/pmrmilestones.pdf
23. Accreditation Council for Graduate Medical Education. Sports Medicine Milestones. Accreditation Council for Graduate Medical Education; 2022. Accessed January 26, 2024. https://www.acgme.org/globalassets/pdfs/milestones/sportsmedicinemilestones.pdf
24. Kissin EY, Nishio J, Yang M, et al. Self-directed learning of basic musculoskeletal ultrasound among rheumatologists in the United States. Arthritis Care Res (Hoboken) 2010;62(2):155–160.
25. Molina Collada J, Mayordomo L; Grupo de Trabajo de Ecografía de la Sociedad Española de Reumatología. Perceived quality of training in musculoskeletal ultrasound in Spanish rheumatology services. Reumatol Clin (Engl Ed) 2022;18(6):349–354.
26. Wu M, Peng L, Donroe JH, et al. Musculoskeletal ultrasound imaging training, use, and knowledge among rheumatologists in China. Clin Rheumatol 2021;40(1):321–330.
27. Kellerman SE, Herold J. Physician response to surveys. A review of the literature. Am J Prev Med 2001;20(1):61–67.
28. Groves RM, Presser S, Dipko S. The role of topic interest in survey participation decisions. Public Opin Q 2004;68(1):2–31.

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 666–675
DOI 10.1002/acr.25471

American College of Rheumatology
Empowering Rheumatology Professionals

Incidence of Side Effects Associated With Acetaminophen in People Aged 65 Years or More: A Prospective Cohort Study Using Data From the Clinical Practice Research Datalink

Jaspreet Kaur,[1] Georgina Nakafero,[2] Abhishek Abhishek,[1] Christian Mallen,[3] Michael Doherty,[4] and Weiya Zhang[4]

Objective. The main objective of this study is to examine the safety of oral acetaminophen at its therapeutic dose in adults aged ≥65 years.

Methods. This population-based cohort study used the Clinical Practice Research Datalink-Gold data. Participants were aged ≥65 years registered with a UK general practice for at least 12 months between 1998 and 2018. Acetaminophen exposure was defined as at least two acetaminophen prescriptions within six months of the first acetaminophen prescription, the first prescription date being the index date. Acetaminophen nonexposure was defined as the absence of two acetaminophen prescriptions within six months over the study period. We calculated propensity score (PS) for acetaminophen prescription and undertook inverse probability treatment weighting using PS and PS-matched analyses to account for confounding. Missing data were handled using multiple imputation. The adjusted hazard ratio (aHR) and 95% confidence interval (95% CI) were calculated using the Cox proportional hazards regression model.

Results. In total, 180,483 acetaminophen exposed and 402,478 unexposed participants were included in this study. Acetaminophen exposure was associated with an increased risk of perforation or ulceration or bleeding (aHR 1.24; 95% CI 1.16–1.34), uncomplicated peptic ulcers (aHR 1.20; 95% CI 1.10–1.31), lower gastrointestinal bleeding (aHR 1.36; 95% CI 1.29–1.46), heart failure (aHR 1.09; 95% CI 1.06–1.13), hypertension (aHR 1.07; 95% CI 1.04–1.11), and chronic kidney disease (aHR 1.19; 95% CI 1.13–1.24).

Conclusion. Despite its perceived safety, acetaminophen is associated with several serious complications. Given its minimal analgesic effectiveness, acetaminophen as the first-line oral analgesic option for long-term conditions in older people requires careful reconsideration.

INTRODUCTION

Almost all clinical guidelines advocate acetaminophen as the first-line oral pharmacological treatment for pain due to osteoarthritis (OA), mainly because of its perceived safety over other oral analgesics.[1–4] However, recent studies have raised concerns that acetaminophen may be not as safe as previously thought.[5–9]

Acetaminophen can cause cyclooxygenase (COX)–dependent side effects analogous to those of nonsteroidal anti-inflammatory drugs (NSAIDs).[5,10,11] The National Institute for Health and Care Excellence (NICE) in the UK highlighted this concern over the safety of acetaminophen in the 2014,[1] updated guidance on management of OA and no longer recommends acetaminophen as a regular treatment in the 2022 update.[14] This is related to both its nonclinically meaningful benefit, which had been confirmed earlier in a meta-analysis (MA) and network MA,[12,13] and its potential harms.[14]

Ideally, randomized controlled trials (RCTs) are required to provide evidence of drug efficacy and safety. However, RCTs

Supported by the University of Nottingham and received infrastructure support from the Pain Centre Versus Arthritis (United Kingdom).

[1]Jaspreet Kaur, BAMS, MPH, PG Diploma, PhD, Abhishek Abhishek, MBBS, MD (Medicine), FRCP (UK), PhD: University of Nottingham, Nottingham, United Kingdom; [2]Georgina Nakafero, BScN, MPH, PhD: University of Nottingham and National Institute for Health and Care Research Biological Research Centre, Nottingham, United Kingdom; [3]Christian Mallen, BMedSci, BMBS, MMedSci, MPhil, PhD, FRCGP, FFCI, FFPH: Keele University, Keele, United Kingdom; [4]Michael Doherty, MA, MD, FRCP, Weiya Zhang, BMed, MMed, PhD: University of Nottingham and Pain Centre Versus Arthritis, Nottingham, United Kingdom.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25471).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25471.

Address correspondence via email to Jaspreet Kaur, BAMS, MPH, PG Diploma, PhD, at drjaspreet@live.co.uk or jaspreet.kaur1@nottingham.ac.uk.

Submitted for publication June 10, 2024; accepted in revised form November 12, 2024.

SIGNIFICANCE & INNOVATIONS

- Acetaminophen is a relatively weak analgesic, but largely because of its perceived safety, it has been recommended as the first-line oral analgesic, especially in older people. However, this study shows a significant incidence of gastrointestinal, cardiovascular, and renal side effects in older people, who are prescribed acetaminophen repeatedly in the United Kingdom.
- Although the incidence may be lower, the side effect profile aligns with that of nonsteroidal anti-inflammatory drugs and cyclooxygenase 2 (COX-2) inhibitors, which reflects the now-recognized COX inhibitory effect of acetaminophen.
- These findings support reconsideration of acetaminophen as an oral analgesic by guideline development groups who currently recommend its repeated prescription for long-term conditions such as osteoarthritis.

are not an optimal design to evaluate the safety of acetaminophen because of ethical concerns and cost implications that preclude the recruitment of enough people in an adequately powered study of sufficient duration. Therefore, the evidence concerning the safety of acetaminophen at its therapeutic dose primarily comes from postmarketing observational studies.[5,8,10,15,16] Previous observational studies on acetaminophen have encountered methodologic challenges, such as channeling bias. This bias occurs when individuals, particularly older adults at higher risk of gastrointestinal and cardiovascular adverse events, are less likely to be given NSAIDs but are more likely to receive acetaminophen.[5]

A previous study that used propensity score (PS) matching, in which the researchers compared the safety profiles of topical NSAIDs with acetaminophen and oral NSAIDs in people with knee or hip OA, revealed that topical NSAIDs exhibited a better safety profile compared with acetaminophen or oral NSAIDS.[17] However, that study did not include a comparison with no analgesics. It is still unclear whether individuals prescribed acetaminophen are at increased risk of developing gastrointestinal, cardiovascular, and renal adverse events compared with those not taking any analgesics.

To address this gap and mitigate channeling bias, we conducted this cohort study comparing acetaminophen exposure versus non-exposure to any analgesics for major adverse events in the general population, as well as in people with OA as a subgroup analysis using the large UK Clinical Practice Research Datalink. The objectives of the study were 1) to examine the incidence of perforation or ulceration or bleeding (PUB), uncomplicated peptic ulcers, lower gastrointestinal (GI) bleeding, heart failure, myocardial infarction, hypertension, and chronic renal failure in people prescribed acetaminophen in the general population compared with people unexposed to analgesics; and 2) to examine the dose-response relationship between acetaminophen prescription and specific adverse events.

PATIENTS AND METHODS

Study design. This was a population-based cohort study, conducted with data from the Clinical Practice Research Datalink (CPRD) GOLD,[18] where acetamiophen exposed participants were compared to acetaminophen unexposed for the incidence of major adverse events, including GI, cardiovascular (CV), and renal adverse events. The CPRD is one of the largest health care databases and has been demonstrated to be a reliable resource for research.[19–21] It had collected anonymized patient data from 736 general practices, covering 17 million UK residents as of January 2018. Data were entered electronically either during a consultation with a general practitioner or through communication with other health professionals. Details of the patient demographics, medications, and diagnoses were recorded.[22] The CPRD broadly represents the UK population in terms of age, gender, ethnicity, and body mass index (BMI).[23] Although not explicitly designed for surveys or studies, such standard health data are cost-effective for research purposes, allowing for examination of the effectiveness and safety of interventions in the real world.[22]

Participants. In the United Kingdom, people aged ≥65 years are eligible for free prescriptions from their general practitioner, which permitted us to examine the safety of acetaminophen in older people in the United Kingdom. We included older adults aged ≥65 years at index date—the date of the first acetaminophen prescription between January 1, 1998, and January 1, 2018, who had been registered for at least 12 months with a general practice. The practices included were deemed up to standard by CPRD, i.e, contributing comprehensive, continuous, and complete high-quality data for research purposes. We excluded participants with a diagnosis of the outcome of interest before the start of follow-up (details given in exposure section below). Consent from individuals involved in this study was not sought because this was database research. The Independent Scientific Advisory Committee approved this cohort study with the protocol reference 19_131R.

Exposure. Participants who were issued at least two acetaminophen prescriptions within six months, and not in combination with other analgesics such as codeine, were defined as exposed to acetaminophen to exclude occasional oral intake of acetaminophen for other acute conditions such as headache and influenza. The date of the first of the two prescriptions was assigned as the index date. All the records were obtained using the product codes (Supplementary Table S1). To avoid prevalent cases, we included only those who did not have an acetaminophen prescription in the 12 months before the index date.[24] A

landmark analysis was applied as a simple approach to minimize immortal time bias—a potential confounding that could influence the results when follow-up is delayed one way or another.[25] Participants were followed from the landmark date (i.e, the date 12 months after the index date) to avoid the chance of counting outcomes during the exposure period. This would also give a 12-month exposure window from the index date to the landmark date to accumulate repeated acetaminophen doses for the dose-response analysis. The follow-up stopped at the earliest event, study end date (January 1, 2018), date of death, or transfer out of practices, whichever came first.

Comparators. The control group (unexposed) included participants aged ≥65 years with less than two acetaminophen prescriptions within six months during the study period when exposure could be accrued. The controls were individually matched to acetaminophen exposed participants by year of birth, sex, and general practice on a 1:n (as available) basis, then further matched by PS on a 1:1 basis. The controls were assigned the index date and landmark date of their matched acetaminophen-exposed participant.

Outcomes. The outcomes of interest were the incident diagnosis of 1) GI conditions, specifically PUB, uncomplicated peptic ulcers (ulcers without bleeding or perforation), and lower GI bleeding; 2) CV outcomes, specifically hypertension, myocardial infarction, or heart failure; and 3) chronic renal failure. Patients with these outcomes were identified using relevant Read codes or Med codes recorded in the CPRD records. These Med codes were similar to those used in other studies to extract data on GI, CV, and renal events and were updated using the medical dictionary of the CPRD data set.[5,21]

Covariates. We shortlisted potential covariates such as age, sex, Charlson comorbidity index (CCI) (Supplementary Material: Covariates), opioids, NSAIDS, coxibs, aspirin, H2-receptor blockers, proton pump inhibitors, dipyridamole, clopidogrel and lifestyle factors (smoking, alcohol, and BMI) based on their association with acetaminophen and the outcome of interest. These covariates were used to calculate the PS for acetaminophen prescription to account for confounding.[26] Codes for the comorbidities were obtained from the Primary Care Unit, University of Cambridge (https://www.phpc.cam.ac.uk/pcu/cprd_cam/codelists/v11/) and updated using the CPRD code browser. We excluded any comorbidity from the CCI if it was an outcome to avoid including prevalent cases (Supplementary Table S2).

Statistical analysis. Logistic regression was used to calculate the PS for acetaminophen prescription.[27] Standardized differences (*d*) were used to examine the covariate balance between the exposed and unexposed participants. Covariates for which there was an imbalance, defined as $d > 0.1$, were included as additional covariates in the subsequent Cox regression model.[28] The kernel-density plot was used to check the covariate balance graphically. The cumulative survival probability was estimated using the Kaplan-Meier (Nelson-Aalen) survival curve.[29] The proportional hazards assumption was assessed using log-log plots and the Schoenfeld residuals.[30] The Cox regression model calculated the hazard ratio (HR) between the acetaminophen-exposed versus nonexposed participants. Different Cox regression models were constructed: Model 1—age, sex, and general practice (GP) matched; Model 2—age, sex, GP, and PS matched; Model 3—age, sex, GP, and PS-matched analyses adjusted for unbalanced covariates; and Model 4—inverse probability treatment weighting (IPTW) using PS. IPTW using PS was considered the main model given its potential to yield more precise estimates in time to event analyses than the PS-matched method.[31]

We conducted a subgroup analysis restricted to participants with OA, a common condition of older people that is often associated with long-term analgesic regime. In this subgroup analysis, the PS model was recalculated. This ensured that the matched pairs were comparable in this subgroup of people with OA and allowed us to capture the safety profile of acetaminophen in a common long-term condition that often requires regular oral administration of acetaminophen. Statistical significance was considered at $P < 0.05$.

To assess a dose-response relationship between acetaminophen and the outcomes, we calculated the number of acetaminophen prescriptions between index and landmark date. These were categorized as no prescription (reference group), one to two prescriptions, three to four prescriptions, five to six prescriptions, seven to eight prescriptions, and nine or more. To minimize channeling bias, the dose-response analyses were repeated with the individuals prescribed one to two prescriptions as the reference category.

Missing data on BMI, alcohol use, and smoking were handled as a separate category in the PS-matched analyses and by multiple imputation in the IPTW using PS analyses (Supplementary Material: Handling missing data). The data used in this study could be obtained directly from the CPRD upon request because of CPRD licensing rules.

RESULTS

Cohort description. Of 2,789,347 participants (exposure, 697,362 and nonexposure, 2,091,985) matched by age, sex, and practice, 582,961 were included in the IPTW using PS analyses, and 158,048 were successfully PS matched (79,024 acetaminophen exposed, 79,024 unexposed) (Supplementary Figure S1). The mean age ± SD of participants included in the IPTW using PS and PS-matched cohort was 74.88 ± 7.29 years and 75.42 ± 7.51 years, respectively. Most participants self-identified as female (Table 1). The mean ± SD duration of follow-up was 1 ± 4.62 years.

Table 1. Participant characteristics*

	IPTW using PS sample (n = 582,961)[a]			PS-matched sample (n = 158,048)[a]		
	Acetaminophen exposed	Acetaminophen unexposed	*d*	Acetaminophen exposed	Acetaminophen unexposed	*d*
n	180,483	402,478		79,024	79,024	
Continuous variables, mean ± SD						
Age	74.88 ± 7.29	74.99 ± 7.47	−0.01	75.42 ± 7.51	75.41 ± 7.75	0.00
Charlson's comorbidity index	0.43 ± 1.24	0.24 ± 0.93	0.15	0.53 ± 1.45	0.36 ± 1.21	**0.13**
Categorical covariates, n (%)						
Female	114,066 (63.20)	242,175 (60.17)	0.06	48,167 (60.95)	50,421 (63.80)	0.06
OA	55,038 (30.49)	60,349 (14.99)	0.38	28,356 (35.88)	24,391 (30.86)	**0.11**
BMI						
Underweight	3,647 (2.02)	5,505 (1.37)	0.05	2,330 (2.95)	1,625 (2.06)	0.06
Overweight	43,402 (24.02)	73,328 (18.22)	0.14	22,776 (28.82)	20,024 (25.34)	0.08
Obese	22,163 (12.28)	29,131 (7.24)	0.17	12,363 (15.64)	9,761 (12.35)	0.09
Missing	64,148 (35.54)	208,224 (51.74)	−0.33	19,101 (24.17)	25,648 (32.45)	−0.18
Smoking status						
Current	15,138 (8.39)	30,110 (7.48)	0.03	8,875 (11.23)	8,237 (10.42)	0.03
Ex-smoker	44,190 (24.48)	63,636 (15.81)	0.21	21,471 (27.17)	18,347 (23.21)	0.09
Missing	32,468 (17.99)	168,282 (41.81)	−0.54	9,208 (11.65)	14,680 (18.57)	−0.19
Alcohol						
Current drinkers	79,661 (44.14)	144,373 (35.87)	0.17	40,956 (51.82)	38,149 (48.27)	0.07
Past drinkers	4,448 (2.46)	4,767 (1.18)	0.10	2,180 (2.76)	1,493 (1.89)	0.06
Missing	57,999 (32.14)	198,151 (49.23)	−0.35	16,978 (21.48)	22,904 (28.98)	−0.17
Other medication						
Opioids	178,328 (69.87)	76,896 (30.13)	2.76	76,878 (97.27)	76,878 (97.27)	0.00
NSAIDS	55,559 (30.78)	48,061 (11.94)	0.47	26,635 (33.70)	24,385 (30.85)	0.06
Coxibs	10,077 (5.58)	5,965 (1.48)	0.22	5,208 (6.59)	3,621 (4.58)	0.09
Clopidogrel	10,183 (5.64)	7,658 (1.90)	0.20	4,073 (5.15)	2,983 (3.77)	0.07
Dipyridamole	5,255 (2.91)	4,947 (1.23)	0.12	2,771 (3.51)	1,835 (2.32)	0.07
H2-receptor blockers	18,745 (10.39)	17,714 (4.40)	0.23	10,914 (13.81)	8,213 (10.39)	**0.10**
Proton pump inhibitors	63,710 (35.30)	49,621 (12.33)	0.56	25,060 (31.71)	22,250 (28.15)	0.07
Aspirin	72,064 (39.93)	78,523 (19.51)	0.46	31,257 (39.55)	26,428 (33.44)	**0.13**

* Significant values are in bold. BMI, body mass index; *d*, standardized difference; IPTW, inverse probability treatment weighting; NSAID, non-steroidal anti-inflammatory drug; OA, osteoarthritis; PS, propensity score.
[a] After age, sex, and general practice matching.

The Kaplan-Meier curves for acetaminophen exposed compared with unexposed for PUB, uncomplicated peptic ulcers, lower GI bleeding, heart failure, myocardial infarction, hypertension, and chronic renal disease before PS matching are given in Figure 1A and C, Figure 2A and C, Figure 3A and C, and Supplementary Figure S2A, respectively. The results after PS matching are given in Figure 1B and D, Figure 2B and D, Figure 3B and D, and Supplementary Figure S2B, respectively. These estimates indicated that the cumulative hazards for PUB, uncomplicated peptic ulcers, lower GI bleeding, heart failure, myocardial infarction, hypertension, and chronic renal disease were higher in the acetaminophen-exposed group compared with the unexposed group.

Incidence of GI, CV, and renal events. There was an increased incidence of PUB, uncomplicated peptic ulcers, lower GI bleeding, heart failure, hypertension, and chronic renal failure in the acetaminophen-exposed participants compared with the unexposed participants with adjusted HR (aHR) (95% confidence interval [95% CI]) after IPTW using PS of 1.24 (1.16–1.34), 1.20 (1.10–1.31), 1.36 (1.29–1.46), 1.09 (1.06–1.13), 1.07 (1.04–1.11), and 1.19 (1.13–1.24), respectively. Similar results were observed from the PS-matched analyses (Table 2).

Dose response. The association among developing PUB, uncomplicated peptic ulcers, and chronic renal failure increased with the number of acetaminophen prescriptions (Supplementary Table S3). A similar trend was observed when the analysis was restricted to the acetaminophen-exposed–only group, with *P* for trend <0.01 (Supplementary Table S4).

Subgroup analysis. There were 115,387 participants with OA, of whom 48,812 were matched on PS (24,406 acetaminophen exposed and 24,406 unexposed). In this population, exposure to acetaminophen was associated with increased incidence of lower GI bleeding, hypertension, and chronic renal failure with aHR (95% CI) 1.20 (1.09–1.33), 1.06 (1.00–1.13), and 1.15 (1.09–1.22), respectively (Table 3).

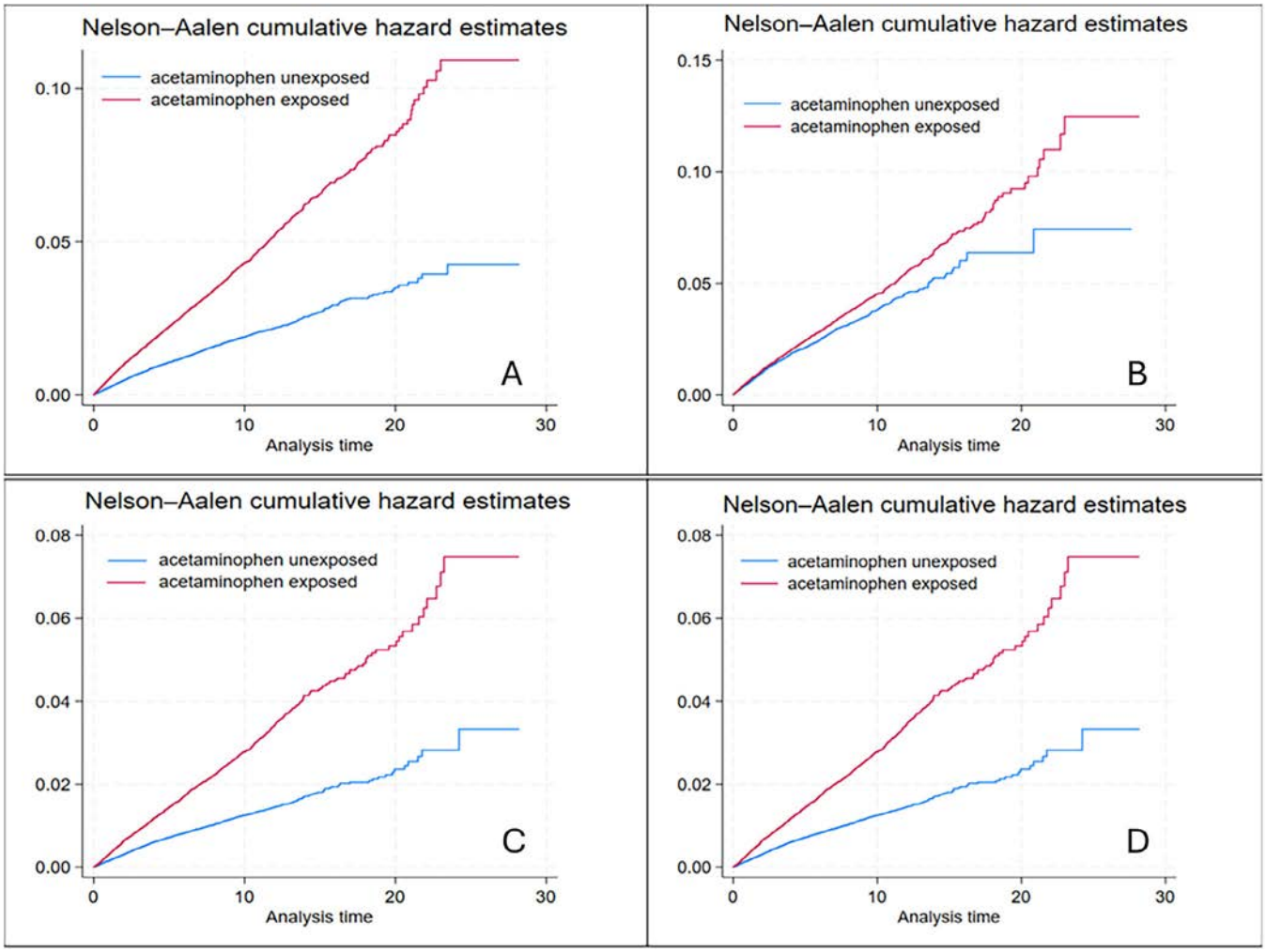


Figure 1. Cumulative hazard estimates before and after PS matching for perforation or ulceration or bleeding (A and B) and uncomplicated peptic ulcers (C and D) (after age, sex, and general practice matching). The red line represents that acetaminophen exposure has higher hazards for perforation or ulceration or bleeding and uncomplicated peptic ulcers than the unexposed group, represented by the blue line. PS, propensity score.

DISCUSSION

In this large study of 180,483 acetaminophen exposed participants and 402,478 unexposed participants aged 65 years and older in the UK primary care population, we found that acetaminophen exposure was associated with an increased incidence of PUB, uncomplicated peptic ulcers, lower GI bleed, heart failure, hypertension, and chronic renal failure. A dose-response relationship was observed for PUB, uncomplicated peptic ulcers, and chronic renal failure. The robustness of these associations was supported by 1) the observed risk for PUB, uncomplicated peptic ulcers, and chronic renal failure across different models; 2) a similar dose-response relationship when the analyses were restricted only to the acetaminophen-exposed group; 3) an increased incidence of lower GI bleeding, hypertension, and chronic renal diseases observed in a subgroup analysis restricted to people with OA—a common long-term painful condition often requiring regular analgesics aligned with that found in the overall cohort.

The findings of our study are consistent with previous observational studies that have reported an association between acetaminophen intake and the risk of GI complications and hypertension.[8,15,17,32–35] According to an experimental study, acetaminophen exerts an inhibitory effect on peripheral COX enzymes, suggesting that it could be a possible mechanism for the GI bleeding associated with its prescription.[7] Furthermore, the significant dose-response relationship of an increased risk associated with acetaminophen exposure and PUB, uncomplicated ulcer, and chronic kidney disease according to the number of prescriptions also aligns with previous observational studies.[5,16,36,37]

The majority of acetaminophen RCTs have not found any major adverse effects,[38–40] apart from one that reported a drop in hemoglobin of ≥1 g/dL over 13 weeks, presumed to be due to GI bleeding, in 20% of participants with knee pain taking acetaminophen 3 g/day.[41] This is because the RCTs were primarily designed for efficacy rather than adverse events, solely reported

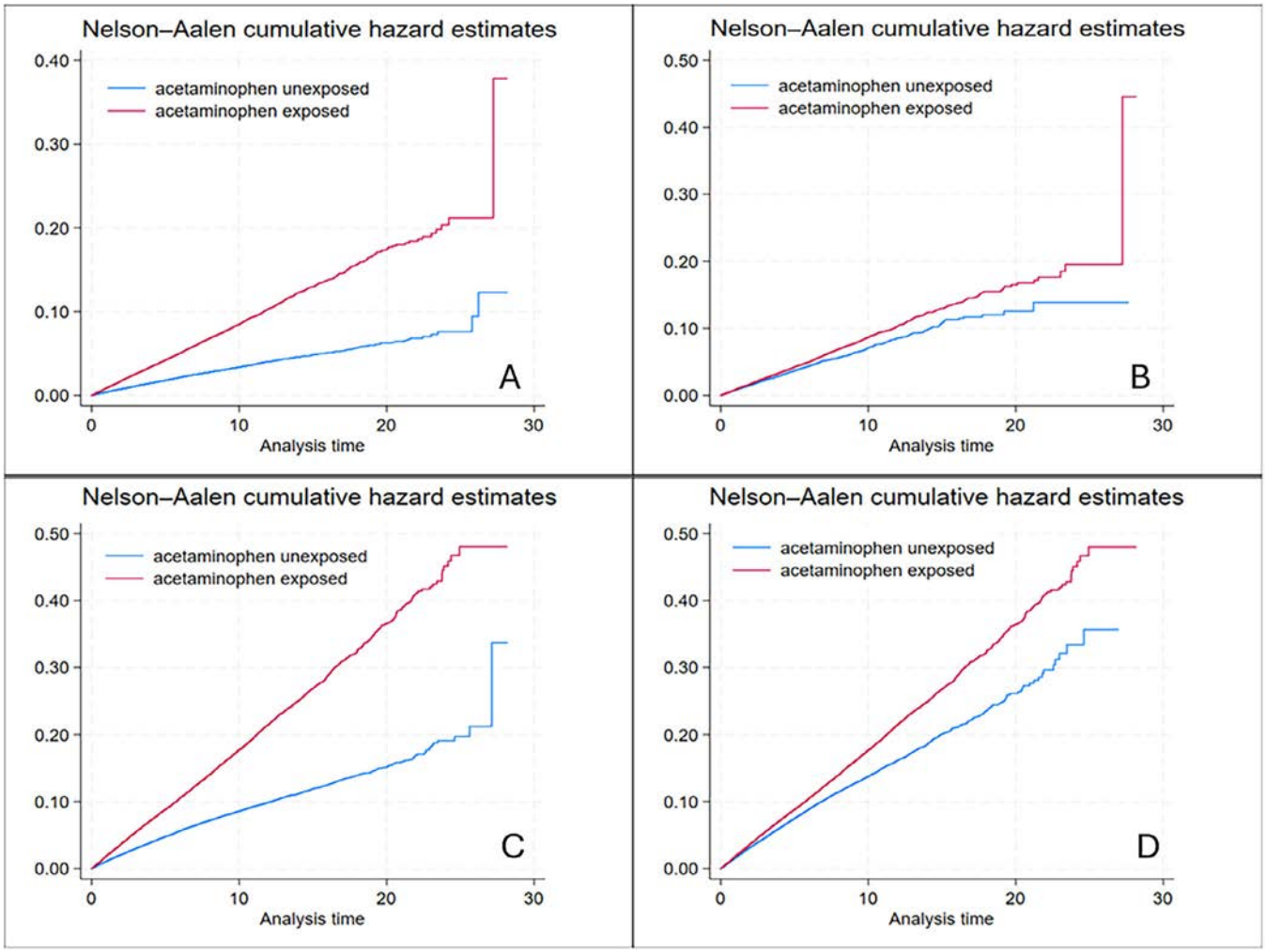


Figure 2. Cumulative hazard estimates for lower gastrointestinal bleeding and heart failure before PS matching (A and C, respectively) and after PS matching (B and D, respectively) (after age, sex, and general practice matching). The red line represents the acetaminophen exposure group and shows higher hazards for lower gastrointestinal bleeding and heart failure than the unexposed group, represented by the blue line. PS, propensity score. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25471/abstract.

short-term effects, were less powered, and recruited healthier and younger participants.

There is limited experimental evidence to support the impact of acetaminophen on GI, CV, or renal events. For example, prolonged acetaminophen ingestion might inhibit prostacyclin synthesis in humans, resulting in GI lesions and bleeding.[42–44] Grèen et al[42] suggested that acetaminophen could disadvantage people suffering from conditions in which prostacyclin-mediated vascular defense mechanisms are pronounced, such as myocardial infarction, deep vein thrombosis, and after surgery. The reason could be that oral administration of 500 mg acetaminophen decreases urinary excretion of 2-3-dinor 6 keto prostaglandin F1alpha, a stable inactive metabolite of major endothelium-derived COX-2 prostacyclin,[42] and can cause marked reduction in prostacyclin synthesis for a minimum of six to eight hours without affecting thromboxane production. Acetaminophen is a major metabolite of phenacetin, which has been associated with hepatotoxicity and renal damage, but the mechanism of renal toxicity due to acetaminophen is still debatable. Furthermore, Lorz et al[45] suggested that acute tubular necrosis might be responsible for renal impairment in people taking acetaminophen long-term. The tubular cells undergo endoplasmic reticulum (ER) stress preceding growth arrest and DNA damage–inducible protein 153 stimulation and alteration to the nucleus in addition to caspase-12 cleavage. Therefore, acetaminophen may cause the induction of caspase-mediated cell death, indicating its nephrotoxic potential, and ER stress could be recognized as a therapeutic target in nephrotoxicity.

The IPTW using PS and PS matching can only control the known confounding factors and cannot address unmeasured or unknown confounders. Therefore, our results are prone to potential confounding bias due to unmeasured/unknown confounders.[21,46]

Only 27% of the study participants were included in the PS-matched analyses, leading to a significant reduction in our study

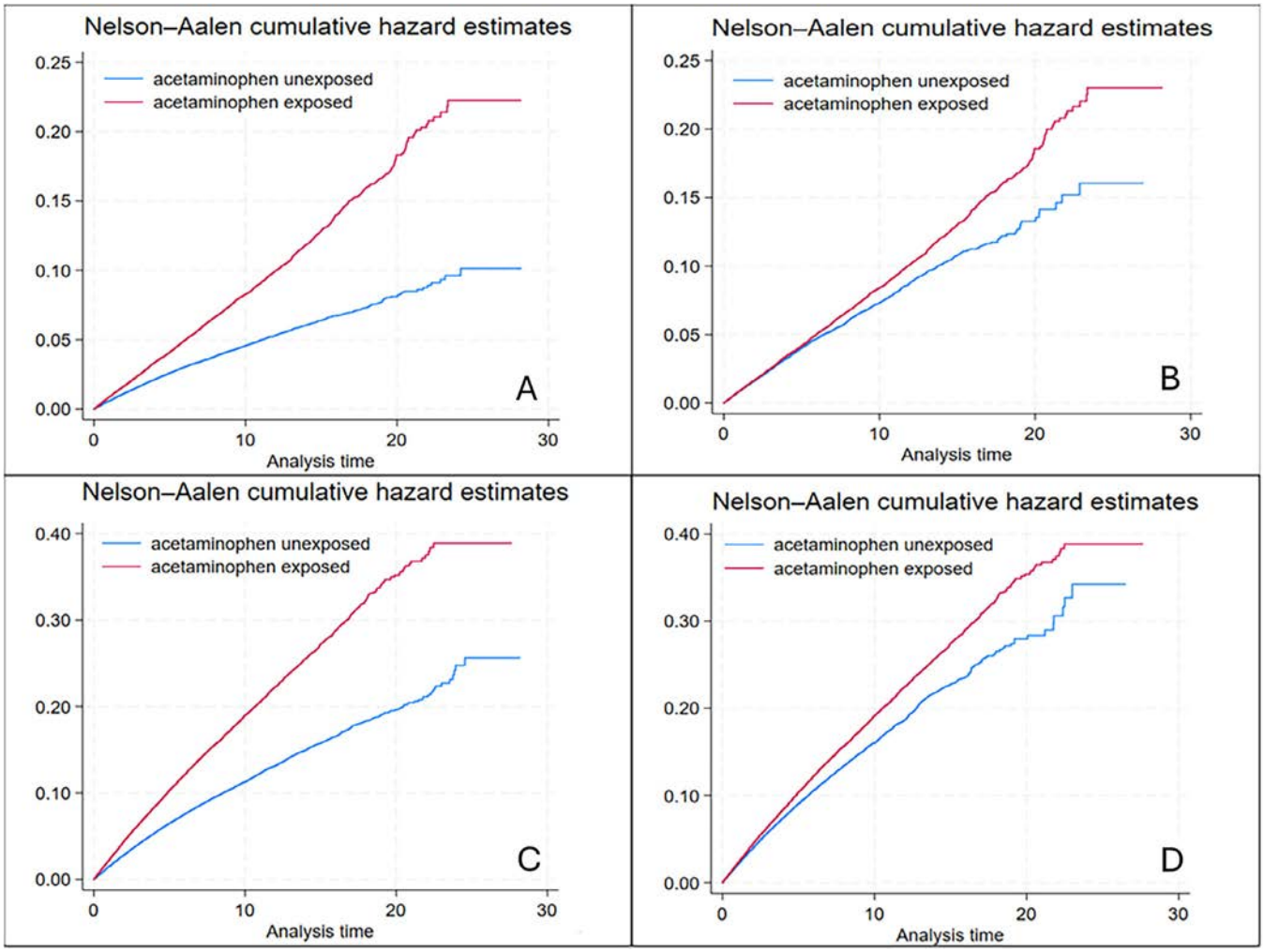


Figure 3. Cumulative hazard estimates for myocardial infarction and hypertension PS matching (A and C, respectively) and after PS matching (B and D, respectively) (after age, sex and general practice matching). The red line represents the acetaminophen exposure group and shows higher hazards for myocardial infarction and hypertension than the unexposed group, represented by the blue line. PS, propensity score. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25471/abstract.

sample size. However, results from the PS-matched analyses were consistent with those from the IPTW using PS, which used all study participants exposed to acetaminophen, providing internal validity of our findings and enhancing their generalizability to older people prescribed acetaminophen.

A significant caveat to the study is that there is no provision for recording over-the-counter prescriptions in the CPRD. This limitation was a reason to restrict the study to people aged ≥65 years, who were eligible for free prescriptions and therefore were less likely to purchase acetaminophen independently. Other reasons for selecting this age group include the fact that older people are at higher risk of GI, CV, and renal adverse events, making them more likely to be prescribed acetaminophen than oral NSAIDs and opioids. Despite this, over-the-counter (OTC) usage might have affected both the exposure and nonexposure groups. It was assumed that the distribution would be random for other OTC analgesics but slightly more toward the nonexposure group, as those prescribed acetaminophen may be less likely to purchase it independently. Therefore, any such imbalance was expected to underestimate rather than overestimate the HR. The findings might be more applicable to settings where acetaminophen is predominantly prescribed as a medication. In populations where OTC remedies are more common across ages, the observed associations might differ.

Furthermore, acetaminophen is often taken episodically and for multiple reasons, making it difficult to define people exposed to acetaminophen and those not exposed to acteminophen. In this study, minimal acetaminophen exposure was defined as two or more prescriptions within six months of the first acetaminophen prescription. However, the subsequent intake of acetaminophen over time was unknown. The immortal time bias could occur not only within the exposure window of the initial 12-month period, but also throughout the entire follow-up period. Therefore, a time-varying exposure analysis is necessary to account for

Table 2. HRs and 95% CIs for incidence of gastrointestinal, cardiovascular, and renal outcomes in the acetaminophen-exposed group vs the nonexposed group*

Outcomes	Acetaminophen exposure	Event rate/ 1,000 person-years (IPTW using PS sample)	Event rate/ 1,000 person-years (PS-matched sample)	Model 1, HR (95% CI)	Model 2, HR (95% CI)	Model 3, HR (95% CI)	Model 4, HR (95% CI)
PUB	No	5.51	11.54	**1.00**	**1.00**	**1.00**	**1.00**
	Yes	12.19	12.63	**2.21 (2.10–2.32)**	**1.10 (1.01–1.19)**	**1.06 (1.00–1.16)**	**1.24 (1.16–1.34)**
Uncomplicated peptic ulcers	No	3.65	7.72	**1.00**	**1.00**	**1.00**	**1.00**
	Yes	7.84	8.38	**2.15 (2.02–2.29)**	**1.09 (1.00–1.21)**	1.04 (0.95–1.16)	**1.20 (1.10–1.31)**
Lower GI bleed	No	9.53	19.87	**1.00**	**1.00**	**1.00**	**1.00**
	Yes	23.36	23.23	**2.45 (2.36–2.54)**	**1.20 (1.13–1.28)**	**1.15 (1.09–1.23)**	**1.36 (1.29–1.45)**
Heart failure	No	24.67	40.94	**1.00**	**1.00**		**1.00**
	Yes	48.97	49.52	**1.98 (1.93–2.03)**	**1.24 (1.20–1.27)**	NA	**1.09 (1.06–1.13)**
Myocardial infarction	No	13.23	23.00	**1.00**	**1.00**		**1.00**
	Yes	22.95	23.13	**1.73 (1.67–1.79)**	**1.04 (1.00–1.09)**	NA	0.99 (0.94–1.04)
Hypertension	No	33.20	50.16	**1.00**	**1.00**		**1.00**
	Yes	54.23	54.57	**1.62 (1.59–1.66)**	**1.10 (1.07–1.13)**	NA	**1.07 (1.04–1.11)**
Chronic renal failure	No	16.24	30.77	**1.00**	**1.00**		**1.00**
	Yes	37.08	40.90	**2.29 (2.25–2.33)**	**1.31 (1.27–1.35)**	**1.22 (1.18–1.25)**	**1.19 (1.13–1.24)**

* Significant results are in bold. Model 1, age, gender, and practice matched; model 2, age, gender, practice and PS matched; model 3, age, gender, practice and PS matched with further adjustments for the variables that were still not balanced after PS matching, ie, Charlson comorbidity index, OA, H2-receptor blockers, aspirin for GI and Charlson comorbidity index, ex-smokers for chronic renal failure; model 4, IPTW method using PS. GI, gastrointestinal; HR, hazard ratio; IPTW, inverse probability treatment weighting; NA, not applicable; OA, osteoarthritis; PS, propensity score; PUB, perforation or ulceration or bleeding; 95% CI, 95% confidence interval.

dynamic changes in oral consumption of acetaminophen, to allow the counting of events according to exposure and nonexposure after the landmark date, and to better reflect the episodic taking of acetaminophen and its association with adverse effects, such as GI bleeding.

This population-based cohort study, in which the likelihood of confounding by indication was minimized using PS methods, has produced robust evidence concerning the safety of acetaminophen in older adults. Although the incidence of acetaminophen side effects may be lower than that of NSAIDs and COX-2 inhibitors, their side effect profiles are similar, which reflects the now-recognized COX inhibitory effect of acetaminophen.[39] These data further challenge whether acetaminophen should be retained as the first-line oral analgesic, especially in older people for common chronic painful conditions, given its nonclinically meaningful benefits and potential harms, and support the recent recommendation by NICE to not prescribe acetaminophen for OA.[14] A study in which acetaminophen prescription is modeled as a time-varying exposure should be undertaken to confirm these findings.

This study provides the most recent evidence regarding the risk of important adverse events associated with oral administration of acetaminophen in the general population, as well as people with OA aged ≥65 years. Given the low analgesic benefit of

Table 3. Subgroup analysis in participants with OA*

Outcomes	Model 1, HR (95% CI)	Model 2, HR (95% CI)	Model 3, HR (95% CI)	Model 4, HR (95% CI)
PUB	**1.22 (1.10–1.35)**	1.08 (0.94–1.26)	1.07 (0.92–1.24)	1.13 (0.99–1.29)
Uncomplicated peptic ulcers	**1.15 (1.01–1.30)**	0.97 (0.81–1.16)	0.95 (0.79–1.14)	1.04 (0.89–1.22)
Lower GI bleed	**1.29 (1.19–1.39)**	1.09 (0.98–1.22)	1.05 (0.94–1.18)	**1.20 (1.09–1.33)**
Heart failure	**1.08 (1.03–1.14)**	**1.06 (1.00–1.11)**	**1.07 (1.02–1.13)**	0.98 (0.92–1.04)
Myocardial infarction	1.00 (0.93–1.07)	0.96 (0.90–1.03)	0.96 (0.88–1.03)	0.82 (0.75–1.04)
Hypertension	0.96 (0.91–1.01)	0.97 (0.93–1.02)	0.99 (0.94–1.05)	**1.06 (1.00–1.13)**
Chronic renal failure	**1.26 (1.22–1.31)**	**1.21 (1.15–1.28)**	**1.16 (1.10–1.22)**	**1.15 (1.09–1.22)**

* The incidence in the nonexposed group was used as the reference; significant results are in bold. Model 1, age, gender, and practice matched; model 2, age, gender, practice and PS matched; model 3, age, gender, practice and PS matched with further adjustments for the variables that were still not balanced after PS matching, ie, Charlson comorbidity index, OA, H2-receptor blockers, aspirin for GI and Charlson comorbidity index, ex-smokers for cardiovascular and renal events; model 4, inverse probability treatment weighting method using PS. GI, gastrointestinal; HR, hazard ratio; OA, osteoarthritis; PS, propensity score; PUB, perforation or ulceration or bleeding; 95% CI, 95% confidence interval.

acetaminophen in OA and its potential harms, existing guidelines recommending acetaminophen as the first-line oral drug treatment for OA require reassessment.

ACKNOWLEDGMENTS

Members of the Public and Patient Involvement (PPI) group from the Department of Academic Rheumatology, University of Nottingham, contributed to the study methodology and were part of the Advisory Group throughout the project. The study findings and modes of their dissemination were discussed and agreed with PPI members.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr Kaur had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Zhang is the guarantor of the project.

Study conception and design. Doherty, and Zhang.

Acquisition of data. Kaur, Nakafero, and Abhishek.

Analysis and interpretation of data. Kaur, Nakafero, Abhishek, Mallen, Doherty, and Zhang.

REFERENCES

1. National Clinical Guideline Centre (UK). Osteoarthritis: Care and Management in Adults. London: National Institute for Health and Care Excellence (UK); 2014.
2. Kloppenburg M, Kroon FP, Blanco FJ, et al. 2018 Update of the EULAR recommendations for the management of hand osteoarthritis. Ann Rheum Dis 2019;78(1):16–24.
3. McAlindon TE, Bannuru R, Sullivan MC, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014;22(3):363–388.
4. Hochberg MC, Altman RD, April KT, et al; American College of Rheumatology. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64(4):465–474.
5. de Vries F, Setakis E, van Staa TP. Concomitant use of ibuprofen and paracetamol and the risk of major clinical safety outcomes. Br J Clin Pharmacol 2010;70(3):429–438.
6. Evans M, Fored CM, Bellocco R, et al. Acetaminophen, aspirin and progression of advanced chronic kidney disease. Nephrol Dial Transplant 2009;24(6):1908–1918.
7. Hinz B, Brune K. Paracetamol and cyclooxygenase inhibition: is there a cause for concern? Ann Rheum Dis 2012;71(1):20–25.
8. Rahme E, Barkun A, Nedjar H, et al. Hospitalizations for upper and lower GI events associated with traditional NSAIDs and acetaminophen among the elderly in Quebec, Canada. Am J Gastroenterol 2008;103(4):872–882.
9. Roberts E, Nunes VD, Buckner S, et al. Paracetamol: not as safe as we thought? A systematic literature review of observational studies. Ann Rheum Dis 2016;75(3):552–559.
10. Rahme E, Marentette MA, Kong SX, et al. Use of NSAIDs, COX-2 inhibitors, and acetaminophen and associated coprescriptions of gastroprotective agents in an elderly population. Arthritis Rheum 2002;47(6):595–602.
11. Rodríguez LA, Hernández-Díaz S. Relative risk of upper gastrointestinal complications among users of acetaminophen and nonsteroidal anti-inflammatory drugs. Epidemiology 2001;12(5):570–576.
12. Machado GC, Maher CG, Ferreira PH, et al. Efficacy and safety of paracetamol for spinal pain and osteoarthritis: systematic review and meta-analysis of randomised placebo controlled trials. BMJ 2015; 350:h1225.
13. da Costa BR, Reichenbach S, Keller N, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. Lancet 2017; 390(10090):e21–e33.
14. National Institute for Health and Care Excellence (NICE). Osteoarthritis in over 16s: diagnosis and management (NICE guideline [NG226]). London: NICE; 2022.
15. Chan AT, Manson JE, Albert CM, et al. Nonsteroidal antiinflammatory drugs, acetaminophen, and the risk of cardiovascular events. Circulation 2006;113(12):1578–1587.
16. Curhan GC, Knight EL, Rosner B, et al. Lifetime nonnarcotic analgesic use and decline in renal function in women. Arch Intern Med 2004; 164(14):1519–1524.
17. Zeng C, Doherty M, Persson MS, et al. Comparative efficacy and safety of acetaminophen, topical and oral non-steroidal anti-inflammatory drugs for knee osteoarthritis: evidence from a network meta-analysis of randomized controlled trials and real-world data. Osteoarthritis Cartilage 2021;29(9):1242–1251.
18. Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol 2015; 44(3):827–836.
19. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;69(1):4–14.
20. Jick SS, Kaye JA, Vasilakis-Scaramozza C, et al. Validity of the General Practice Research Database. Pharmacotherapy 2003;23(5): 686–689.
21. Weinstein RB, Ryan P, Berlin JA, et al. Channeling in the use of nonprescription paracetamol and ibuprofen in an electronic medical records database: evidence and implications. Drug Saf 2017;40(12): 1279–1292.
22. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). BMJ 2018;363:k3532.
23. Campbell J, Dedman DJ, Eaton SC, et al. Is the CPRD GOLD population comparable to the UK population? Pharmacoepidemiology and Drug Safety 2013; 22(Suppl 1):280–281.
24. Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. Am J Epidemiol 2012;175(4):250–262.
25. Lévesque LE, Hanley JA, Kezouh A, et al. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. BMJ 2010;340:b5087.
26. 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.
27. Geldof T, Popovic D, Van Damme N, et al. Nearest neighbour propensity score matching and bootstrapping for estimating binary patient response in oncology: a Monte Carlo simulation. Sci Rep 2020; 10(1):964.
28. Nguyen TL, Collins GS, Spence J, et al. Double-adjustment in propensity score matching analysis: choosing a threshold for considering residual imbalance. BMC Med Res Methodol 2017;17(1):78.
29. Vinogradova Y, Coupland C, Brindle P. Patients who discontinued statin treatment: a protocol for cohort study using primary care data. BMJ Open 2015;5(10):e008701.
30. Stata Survival Analysis and Epidemiological Tables: Reference Manual. Stata Press; 2003.

31. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. Stat Med 2013;32(16):2837–2849.

32. García Rodríguez LA, Hernández-Díaz S. Risk of uncomplicated peptic ulcer among users of aspirin and nonaspirin nonsteroidal antiinflammatory drugs. Am J Epidemiol 2004;159(1):23–31.

33. González-Pérez A, Sáez ME, Johansson S, et al. Risk factors associated with uncomplicated peptic ulcer and changes in medication use after diagnosis. PLoS One 2014;9(7):e101768.

34. Imhof M, Ohmann C, Hartwig A, et al; Düsuk Study Group. Which peptic ulcers bleed? Results of a case-control study. Scand J Gastroenterol 1997;32(2):131–138.

35. Turtle EJ, Dear JW, Webb DJ. A systematic review of the effect of paracetamol on blood pressure in hypertensive and non-hypertensive subjects. Br J Clin Pharmacol 2013;75(6):1396–1405.

36. García Rodríguez LA, Hernández-Díaz S. The risk of upper gastrointestinal complications associated with nonsteroidal anti-inflammatory drugs, glucocorticoids, acetaminophen, and combinations of these agents. Arthritis Res 2001;3(2):98–101.

37. Kurth T, Glynn RJ, Walker AM, et al. Analgesic use and change in kidney function in apparently healthy men. Am J Kidney Dis 2003;42(2):234–244.

38. Lanza FL, Codispoti JR, Nelson EB. An endoscopic comparison of gastroduodenal injury with over-the-counter doses of ketoprofen and acetaminophen. Am J Gastroenterol 1998;93(7):1051–1054.

39. Hoftiezer JW, O'Laughlin JC, Ivey KJ. Effects of 24 hours of aspirin, Bufferin, paracetamol and placebo on normal human gastroduodenal mucosa. Gut 1982;23(8):692–697.

40. Lanza FL, Collaku A, Liu DJ. Endoscopic comparison of gastroduodenal injury with over-the-counter doses of new fast-dissolving ibuprofen and paracetamol formulations: a randomized, placebo-controlled, 4-way crossover clinical trial. Clin Exp Gastroenterol 2018;11:169–177.

41. Doherty M, Hawkey C, Goulder M, et al. A randomised controlled trial of ibuprofen, paracetamol or a combination tablet of ibuprofen/paracetamol in community-derived people with knee pain. Ann Rheum Dis 2011;70(9):1534–1541.

42. Grèen K, Drvota V, Vesterqvist O. Pronounced reduction of in vivo prostacyclin synthesis in humans by acetaminophen (paracetamol). Prostaglandins 1989;37(3):311–315.

43. Prescott LF. Paracetamol: past, present, and future. Am J Ther 2000;7(2):143–147.

44. Konturek SJ, Obtulowicz W, Sito E, et al. Distribution of prostaglandins in gastric and duodenal mucosa of healthy subjects and duodenal ulcer patients: effects of aspirin and paracetamol. Gut 1981;22(4):283–289.

45. Lorz C, Justo P, Sanz A, et al. Paracetamol-induced renal tubular injury: a role for ER stress. J Am Soc Nephrol 2004;15(2):380–389.

46. Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. J Clin Epidemiol 2017;87:23–34.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 676–684
DOI 10.1002/acr.25473

Not So Patient Friendly: Patient Education Materials in Rheumatology and Internal Medicine Fall Short of Nationally Recommended Readability Benchmarks in the United States

Yazmin Rustomji,[1] Ugochukwu C. Nweke,[2] Sobia Hassan,[1] Usama Ahmad,[1] and Meenakshi Jolly[1]

Objective. Patient education materials (PEMs) can help promote health literacy (HL) among patients. However, their use depends on how easily patients can read and comprehend the information. Several national organizations recommend that text be written at a sixth- to eighth-grade level. Herein, we assess and compare the readability and comprehension (RC) of PEMs for rheumatologic and general medical conditions.

Methods. We used six standardized RC metrics including the well-known Flesch Kincaid Readability Ease and Flesch Kincaid Grade Level to evaluate the RC of PEMs (n = 175) on the American College of Rheumatology (ACR) (n = 86) and the Journal of the American Medical Association (JAMA) (n = 89) websites. Two-sided *t*-tests compared RC between the two resources. $P \leq 0.05$ was considered significant.

Results. On all six standardized metrics used, the mean reading level of all PEMs ranged from high school to college level. For example, the mean ± SD of Simple Measure of Gobbledygook Index was 10.89 ± 1.88, corresponding to a 10th-grade education, and the mean ± SD of Gunning Fog Score was 14.39 ± 2.49, corresponding to a 14th-grade education required to understand the text. JAMA PEMs had significantly more difficult RC levels compared to ACR PEMs based on five of the six indices used ($P < 0.05$).

Conclusion. PEMs available on the ACR and JAMA websites do not align with national organizations' recommendations for RC levels. To enhance patient understanding and promote HL, existing PEMs must be modified in line with these recommendations.

INTRODUCTION

In 2003, the National Center for Education Statistics conducted a national assessment of adult literacy, including a health literacy (HL) assessment with three domains including clinical, prevention, and navigating the health care system. Results revealed that nearly one-third of American adults struggled to comprehend and act on health-related information.[1] HL encompasses the ability to read, understand, and, most importantly, apply health information. Adequate HL is crucial for individuals' ability to make informed health care decisions, prevent illness, and promote wellness.[2]

The National Assessment of Adult Literacy identifies four levels of HL: below basic, basic, intermediate, and proficient. Only 12% of adults in the United States meet a proficient level. In contrast, more than one-third of adults fall into the below basic and basic levels and thus have a limited ability to engage in complex health-related activities. This corresponds to approximately 77 million individuals facing challenges with basic health tasks such as following vaccination guidelines or understanding prescription drug labels.[1,3]

Individual and social factors such as a patient's educational status, cultural norms, and language proficiency, as well as the ability of the media, government, and marketplaces to convey health information, can shape one's HL level.[4] Additionally, health care providers' ability to meaningfully educate patients about their health is crucial for patients to participate in shared decision-making.[5] This is especially important regarding the complex and chronic conditions encountered within rheumatology and internal

Supported by the Brewer Foundation.
[1]Yazmin Rustomji, MD, Sobia Hassan, MD, Usama Ahmad, MD, Meenakshi Jolly, MD, MS: Rush University Medical Center, Chicago, Illinois; [2]Ugochukwu C. Nweke, MD: Rush University Graduate College, Chicago, Illinois.
Drs Rustomji and Nweke 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.25473).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25473.
Address correspondence via email to Yazmin Rustomji, MD, at yazmin_rustomji@rush.edu.
Submitted for publication April 26, 2024; accepted in revised form November 19, 2024.

SIGNIFICANCE & INNOVATIONS

- A high school or college level of education is required to read and comprehend the currently available patient education materials (PEMs) from the American College of Rheumatology (ACR) and Journal of the American Medical Association (JAMA) websites. This exceeds the recommended sixth- to eighth-grade reading level for patient education materials proposed by national organizations.
- JAMA PEMs are more difficult for patients to read and comprehend than ACR PEMs.
- Current PEMs should be revised for the appropriate reading level to reflect the health literacy needs of patients, and new PEMs should be created in line with national recommendations.
- PEMs should state a Readability Ease Score for transparency and accountability.

medicine. Patients must have access to effective patient educational materials (PEMs) that facilitate understanding of the disease and its management and, most importantly, provide guidance on how to use that information to seek preventive and ongoing care. PEMs that are readily available and easy to comprehend are critical tools that allow patients to be proactive about their medical care, potentially leading to decreased health care costs, reduced disability, and improved health outcomes.[6]

Arthritis is the number one cause of work disability among US adults and is associated with significant health care costs.[7] In 2013, the total attributable medical costs and earning losses for osteoarthritis were a staggering $303.5 billion, about 1% of the US Gross Domestic Product.[8] A survey of patients with arthritis found that health care professionals were the primary source of information, followed by PEMs.[9] The study revealed that higher levels of information reception were associated with better medication adherence and greater satisfaction with doctors' support for medication-related concerns.[9] Effective PEMs may help with information reception on the individual level and lead to systemic changes, including lower rates of hospitalizations and lower health care costs.

Current PEMs often contain technical jargon and complex language, hindering readability and comprehension (RC).[10] For example, studies have found that PEMs related to osteoarthritis and cardiovascular diseases, two prevalent chronic conditions, did not meet the recommended readability criteria.[11,12] If PEMs do not align with a patient's RC level, patients will struggle to read the entire material, or may not even attempt to read it. Lengthy PEMs are less likely to be read completely; a survey study regarding nutrition and wellness PEMs found that most respondents (83%) preferred single-page PEMs.[13] Available languages pose another barrier; in the same study, 45% of respondents requested materials to be available in Spanish.[13] The current state of PEMs suggests a lack of accessibility for a significant portion of the patient population. Studies have shown that approximately 30% of the adult White population, 58% of the Black population, and 65% of the adult Hispanic population have basic or below basic HL skills.[14] Even individuals who have completed high school or college may face challenges with their HL skills.[15]

Different organizations have varying recommendations for the readability level of PEMs. The National Institutes of Health (NIH) and the American Medical Association (AMA) recommend that PEMs should not exceed the sixth-grade reading level.[15,16] Meanwhile, the Centers for Disease Control and Prevention (CDC) recommends a readability level below the eighth-grade level.[17] Prior studies on the RC of PEMs in arthritis and other health conditions encountered by primary care providers have shown that many of these materials do not meet this criteria[10,18,19]; however, these were conducted nearly a decade ago.

This study aimed to evaluate and compare the current readability and comprehensibility of rheumatology and internal medicine PEMs available on the websites of the American College of Rheumatology (ACR) and the Journal of the American Medical Association (JAMA) using standardized indices. To account for the differences between national reading level recommendations, our study considers the sixth-grade reading level preferable but considers the seventh-grade level acceptable.

MATERIAL AND METHODS

This was a cross-sectional study. PEMs were identified from the ACR and JAMA Network websites. The ACR and JAMA organizations were selected for this study, as they are nationally recognized organizations broadly covering rheumatology and internal medicine and are freely accessible by both patients and clinicians on the internet. Because it is difficult to generalize what resources different clinics and institutions use, these two organizations were chosen as popular resources for patients and clinicians. Additionally, because clinics and institutions may use a variety of electronic medical record (EMR) systems, both organizations offer open access to PEMs. All PEMS from the ACR website under the "Diseases & Conditions" and "Treatments" sections were analyzed. The remaining sections of "Living with Rheumatic Disease" and "Health Care Team" were excluded because these sections did not specifically contain information for rheumatologic diseases and treatments, or they only contained weblinks to PEMs within the previous two sections. From the JAMA Network website, internal medicine PEMs from January 2021 to July 2022, during which this study was conducted, were analyzed (Supplementary Table 1).

Standardized RC metrics were calculated using a free web-based tool called www.webfx.com. This tool contains readability indices to determine the reading grade level of English-language text. The indices used in this study include the Flesch Kincaid Readability Ease (FKRE), Flesch Kincaid Grade Level (FKGL),

Simple Measure of Gobbledygook Index (SMOG), Gunning Fog Score (GFS), Coleman-Liau index (CLI), and the Automated Readability index (ARI). Each index provides insight into the complexity and comprehensibility of the text based on factors including sentence and word length, syllables per word, and the use of uncommon words.

The FKRE and FKGL formulas were selected as they are the most used readability metrics to assess health care literature.[20] The FKRE rates readability on a scale of 0 to 100, with higher scores indicating greater reading ease. The FKGL estimates the grade level required to comprehend the text in the United States. The SMOG index was selected as it is the recommended reading formula for evaluating health care materials.[21] The SMOG index determines the estimated number of education years needed to comprehend the materials. GFS, CLI, and ARI were also selected because multiple other studies that evaluated the RC of PEMs used these formulas.[20] The GFS calculates the estimated number of formal education years needed to understand the text after one reading. The CLI evaluates the US grade level required to comprehend the text, and the ARI determines the grade level needed to understand the passage. Goal scores for each readability index were set at or below seventh grade (Table 1).

Text from each individual PEM was copied and pasted into the WebFX readability calculator to obtain a score based on the chosen formula. To minimize bias and maintain consistency, only the text relating to the disease or treatment was analyzed. Nonessential content such as author names, affiliations, brand names, proprietary names, references, copyright information, disclaimers, author information, hyperlinks, pictures, and advertisements was excluded. The scores for each PEM were transferred to an Excel spreadsheet for further analysis.

The data analyses were conducted using IBM SPSS statistics. There were no power or sample size calculations as this was a hypothesis-generating, exploratory observational study. Mean and SD were used for descriptive statistics. Range and interquartile range (IQR) were also reviewed. We hypothesized that RC for PEMs would not align with the recommended RC levels for PEMs. An F-test was performed to check for normal versus nonnormal distributions to compare RC metrics between ACR and JAMA PEMs. Nonparametric two-sample *t*-tests were performed. Our hypothesis for the comparisons between ACR and JAMA PEMs was that ACR PEMs would be more complex, resulting in less-favorable RC metrics, given the relative rarity of many rheumatic conditions with multisystem involvement and heterogenous manifestations, compared to conditions routinely seen in general internal medicine. Given the larger member size and resources of internal medicine organizations, we also hypothesized they would be more proficient at producing patient-friendly PEMs. $P \leq 0.05$ was considered significant on two-tailed tests.

RESULTS

In total, 175 PEMs were assessed, 86 PEMs from the ACR and 89 PEMS from the JAMA Network website (Supplementary Table 2). The mean ± SD FKRE of all the PEMs evaluated was 42.05 ± 10.98. FKRE of 30 to 50 equates to "difficult to read, best understood by college graduates." FKRE scores had a range (IQR) of 11.7 to 86.9 (14.1). The mean ± SD FKGL of all PEMs was 12.08 ± 2.25, which correlates to a required 12th-grade level to comprehend material and is characterized as "fairly difficult to read." FKGL scores had a range (IQR) of 4.2 to 18.5 (3.1). The mean ± SD GFS was 14.39 ± 2.48, denoting about 14 years of formal education required to comprehend the text on the first reading, with a range (IQR) of 7.1 to 21.5 (3.2). The mean ± SD SMOG index of all PEMs was 10.89 ± 1.88, indicating at least 10 years of education required to comprehend the writing, with a range (IQR) of 5.0 to 16.2 (2.4). The mean ± SD CLI was 14.55 ± 1.60, which is described as "difficult understanding; the text is best understood by college students," with a range (IQR) of 9.0 to 18.4 (1.9). The mean ± SD ARI of all PEMs was 12.28 ± 2.49, which is described as a "12th grader, and a first-year college student can comprehend the passage," with a range (IQR) of 4.6 to

Table 1. Measurement indices and respective formulas to determine reading grade levels*

Index	Formula	Goal score	Interpretation
Flesch Kincaid Readability Ease	206.835 − 1.015 × (words / sentences) − 84.6 × (syllables / words)	≥70	Scored 0–100, with higher scores indicating greater ease of reading
Flesch Kincaid Grade Level	0.39 × (words / sentences) + 11.8 × (syllables / words) − 15.59	≤7	Grade level to comprehend text
Gunning Fog Score	0.4 × [(words / sentences) + 100 × (complex words / words)]	≤7	No. of years of education required to comprehend text on first reading
Simple Measure of Gobbledygook Index	1.0430 × sqrt(30 × complex words / sentences) + 3.1291	≤7	No. of years of formal education required to comprehend writing
Coleman-Liau Index	5.89 × (characters / words) − 0.3 × (sentences / words) − 15.9	≤7	Grade level required to comprehend text
Automated Readability Index	4.71 × (characters / words) + 0.5 × (words / sentences) − 21.43	≤7	Grade level needed to understand the passage

* sqrt, square root.

19.7 (3.3). The mean ± SD number of sentences used by all PEMs was 30.24 ± 7.90, with a range (IQR) of 3 to 59 (10). Each sentence had a mean ± SD number of words of 18.86 ± 3.92 and a range (IQR) of 11.1 to 30.4 (5.6). The mean ± SD total number of words in all PEMs was 557.38 ± 132.78, with a range (IQR) of 37 to 1,223 (149). Complex words accounted for about one-fifth of total words (mean ± SD: 106.84 ± 37.87), with a range (IQR) of 3 to 348 (46). The average number of syllables per word of all PEMs had a mean ± SD of 1.71 ± 0.12, with a range (IQR) of 1.27 to 2.06 (0.1) (Table 2).

A comparison of the RC of PEMs from the two websites, ACR and JAMA, is shown in Table 3. The mean ± SD FKRE of ACR PEMs was higher than of JAMA PEMs, 45.2 3 ± 8.49 compared to 38.99 ± 12.92 ($P < 0.001$), indicating that ACR PEMs are comparatively easier to read. Similarly, the FKGL mean ± SD for ACR PEMs was 11.07 ± 1.63 and for JAMA PEMs was 13.06 ± 2.34, with JAMA PEMs requiring a higher grade level to understand compared to ACR PEMs ($P < 0.001$). The mean ± SD GFS was also significantly lower for ACR PEMs compared to JAMA PEMs, 13.29 ± 1.94 compared to 15.45 ± 2.49, respectively ($P < 0.001$). The differences in mean ± SD of SMOG indices followed the same pattern, with ACR PEMs having a significantly lower mean ± SD compared to JAMA PEMs, 10.01 ± 1.35 versus 11.73 ± 1.94, respectively ($P < 0.001$). The ARI mean ± SD followed this pattern, with ACR PEMs having a significantly lower mean ± SD of 11.15 ± 1.80 compared to JAMA PEMs with a mean ± SD of 13.27 ± 2.58 ($P < 0.001$). Interestingly, there was no significant difference in CLI mean ± SD between ACR PEMs and JAMA PEMs, with a mean ± SD of 14.57 ± 1.41 and 14.53 ± 1.76, respectively ($P = 0.943$). For the average number of sentences used in the PEMs, there was no significant difference between the mean ± SD of ACR PEMs and JAMA PEMs, 31.06 ± 9.08 versus 29.45 ± 6.52, respectively ($P = 0.412$). ACR PEMs included significantly fewer total words compared to JAMA PEMs, mean ± SD was 504.93 ± 148.12 versus 608.06 ± 91.55, respectively ($P < 0.001$). Although there was a significant difference in mean ± SD of complex words between ACR PEMs and JAMA PEMs, 92.63 ± 39.37 versus 120.81 ± 32.19, respectively ($P < 0.001$), we did not find a significant difference in the percent of complex words ($P = 0.059$). The mean ± SD average number of words per sentence was significantly different between the groups; ACR PEMs had a mean ± SD of 16.55 ± 2.79 whereas JAMA PEMs had a mean ± SD of 21.20 ± 3.55 ($P < 0.001$). For average syllables per word, there was no significant difference between the mean ± SD of the two groups ($P = 0.316$).

Table 2. Readability and comprehension scores of all PEMs reviewed (combined ACR and JAMA)*

Reading Index	Mean	SD	Range	IQR	Interpretation of Mean
FKRE	42.05	10.98	75.2	14.1	Scored from 0 to 100. Higher scores indicate greater reading ease. FKRE of 30–50 is "difficult to read, best understood by college graduates"
FKGL	12.08	2.25	14.3	3.1	FKGL of 12 indicates 12th-grade levels are required to comprehend the material on the page and are "fairly difficult to read"
Gunning Fog Score	14.39	2.48	14.1	3.2	Gunning Fog Score of 13–15 denotes that 13–15 years of formal education is needed to comprehend text on the first reading
SMOG Index	10.89	1.88	11.2	2.4	SMOG index of 10 indicates that 10 years of education is needed to comprehend the text
Coleman-Liau Index	14.55	1.60	9.4	1.9	Coleman-Liau index of 14 indicates "difficult understanding; the text is best understood by college students"
ARI	12.28	2.49	15.1	3.3	ARI of 12 denotes "a 12th grader and a first-year college student can comprehend the passage"
Sentences	30.24	7.90	56	10	N/A
Words	557.38	132.78	1,186	149	N/A
Complex words	106.96	38.48	346	46	N/A
Percentage of complex words	18.98	4.16	25.4	5.2	N/A
Average words in a sentence	18.86	3.92	19.3	5.6	N/A
Average syllables per word	1.72	0.11	0.8	0.1	N/A

* ACR, American College of Rheumatology; ARI, Automated Readability index; FKGL, Flesch Kincaid Grade Level; FKRE, Flesch Kincaid Readability Ease; IQR, interquartile range; JAMA, Journal of the American Medical Association; N/A, not applicable; PEM, patient educational material; SMOG, Simple Measure of Gobbledygook.

Table 3. Comparison between readability of PEMs from ACR and JAMA Network*

Reading Index	ACR		JAMA		Independent sample test
	Mean	SD	Mean	SD	*P* value
Flesch Kincaid Readability Ease	45.23	8.49	38.99	12.92	<0.001
Flesch Kincaid Grade Level	11.07	1.63	13.06	2.34	<0.001
Gunning Fog Score	13.29	1.94	15.45	2.49	<0.001
Simple Measure of Gobbledygook Index	10.01	1.35	11.73	1.94	<0.001
Coleman-Liau index	14.57	1.41	14.53	1.76	0.943
Automated Readability index	11.15	1.80	13.37	2.58	<0.001
Sentences	31.06	9.08	29.45	6.52	0.412
Words	504.93	148.12	608.06	91.55	<0.001
Complex words	92.63	39.37	120.81	32.19	<0.001
Percentage of complex words	18.18	3.42	19.75	4.67	0.059
Average words in a sentence	16.55	2.79	21.10	3.55	<0.001
Average syllables per word	1.71	0.09	1.73	0.12	0.316

* ACR, American College of Rheumatology; JAMA, Journal of American Medical Association.

DISCUSSION

Our study aimed to assess the current RC metrics of PEMs available on the ACR and JAMA Network websites for rheumatology and internal medicine. National guidelines from the NIH, AMA, and CDC recommend that PEMs should be written at a sixth- to eighth-grade reading level to ensure accessibility for a wide range of patients. However, our findings reveal that across all six RC measures used, most PEMs on the ACR and JAMA websites are written at a high school or college level. The FKRE, GFS, CLI, and ARI mean scores across all PEMs evaluated indicate at least a college-level education to read and comprehend the PEMs. Mean FKGL scores across PEMs recommend at least a 12th-grade high school reading level to read and comprehend the PEMs. Although the SMOG index is the most recommended to interpret RC of health-related materials,[21] our study showed a mean SMOG corresponding to the 10th grade; this is still three grade levels above the average national recommendation. The ranges for each RC metric were broad, showing that there were some PEMs within the readability recommendations. However, despite the slight differences in the mean reading levels across the six indices, all mean scores were consistently higher than recommended, and the IQR of each RC metric did not include reading levels below the seventh grade, indicating that the PEMs are truly too difficult for many patients to read, understand, and use for their own health care decisions (Figures 1 and 2).

Contrary to our hypothesis, our study showed that internal medicine PEMs from JAMA were more difficult to read than rheumatology PEMs from ACR. We expected that the RC metrics of ACR PEMs would be more difficult owing to the complexity of rheumatologic conditions, which is heightened by their multiorgan system involvement, heterogeneous manifestations, and treatment with medications that can cause significant toxicity. Additionally, patients may have more familiarity with the more common but still complex general medical conditions such as hypertension and diabetes, through discussions with affected family members and friends. We also speculated that large internal medicine organizations would have more expertise and resource allocation to the development of patient-friendly PEMs, especially related to the common chronic diseases that contribute to significant health burden in the general population. However, our findings dispute our hypotheses. Furthermore, when comparing individual PEMs from ACR and JAMA on the same topic, the ACR PEMs scored higher on FKRE and lower on FKGL, GFS, SMOG index, CLI, and ARI compared to their respective JAMA PEM, corresponding to less difficult RC levels (Supplementary Table 3). One explanation of this difference includes that internal medicine PEMs from the JAMA Network, despite being labeled as "Patients Page," may have been written primarily for academic purposes by content experts but perhaps not by experts in the development of PEMs. As a reputable research journal, JAMA Network focuses on publications and research, which could explain the higher difficulty level of their PEMs.

Patient-friendly PEMs play a crucial role in promoting patient engagement and optimizing HL. Numerous studies have highlighted the link between low HL, poor health care outcomes, and health disparities.[22–24] Low HL marginalizes vulnerable groups and leads to a decreased ability to access health care services, manage chronic diseases, complete health forms, engage in self-care, and communicate effectively with health care providers.[25] Low HL also contributes to higher hospital admission rates, medication nonadherence, increased mortality, increased health care costs, and reduced use of preventive services.[26]

The field of rheumatology has been at the forefront of understanding the needs of its patient population and incorporating those needs into medical management. For example, the general Health Literacy Universal Precautions Toolkit has been adapted for rheumatology clinics[27]; its use has been shown to increase medication compliance and possibly improve disease activity of rheumatoid arthritis.[28] Key aspects of this toolkit include improving written communication. This is where effective PEMs come in. However, despite these studies highlighting the importance of

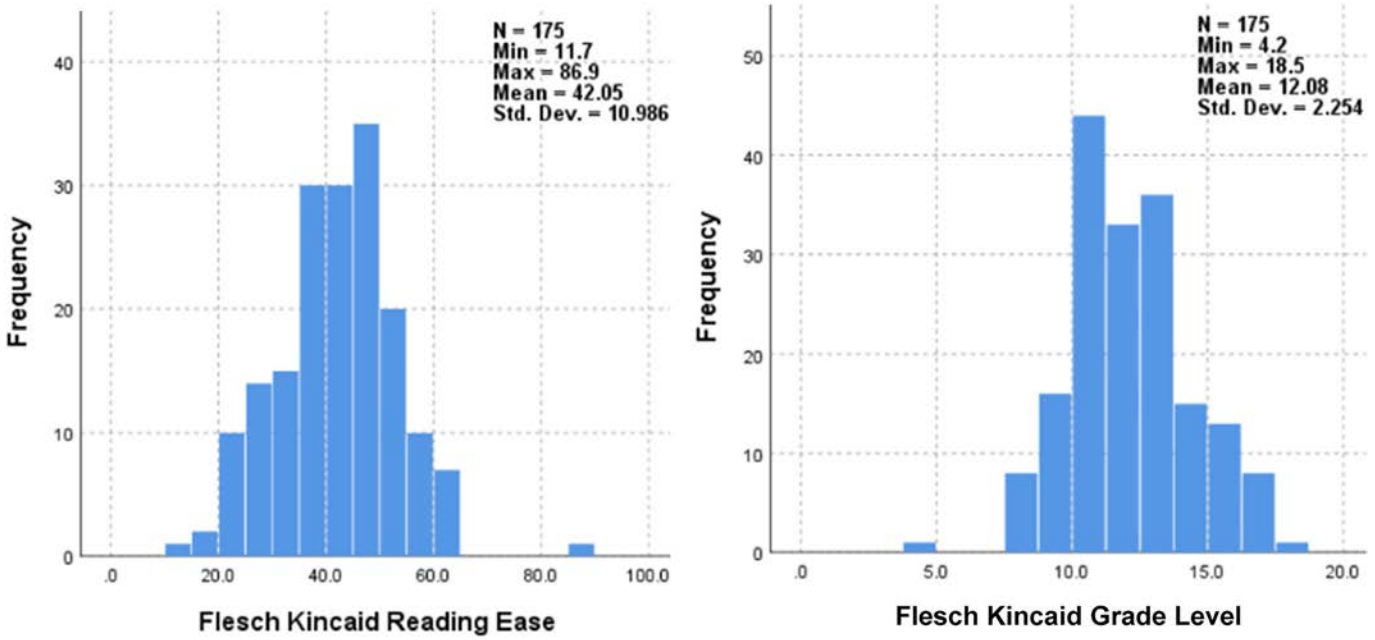


Figure 1. Combined distributions of all patient education materials for the Flesch Kincaid Readability Ease and Flesch Kincaid Grade Level indices.

clearly written communication tailored to the HL needs of patients, our research aligns with previous studies that have shown that many PEMs are difficult for patients to read, understand, and use. A 2013 study comparing online PEMs on osteoarthritis, systemic lupus erythematosus, rheumatoid arthritis, and vasculitis from various resources showed that PEMs from ACR were consistently over the recommended eighth-grade level and were more difficult than PEMs from other national organizations such as the National Institute of Arthritis and Musculoskeletal and Skin Diseases and the Mayo Clinic.[11] A 2021 study showed that only 2.1% of the 2,585 PEMs published in high-impact journals from 1998 to 2018 met the AMA recommendation of sixth-grade reading level, whereas 8.2% met the eighth-grade recommendation of reading level by NIH.[29] In this study, Annals of Rheumatic Disease (ARD) and JAMA repeatedly produced PEMs with reading levels greater than the 11th grade, without significant improvement over the years.[29] Like in our study, the reading grade level for ARD PEMs was lower than JAMA PEMs. A 2019 study, in which patients from Australia, Canada, and the United Kingdom were tested for their knowledge of medications after receiving corresponding PEMs from their rheumatologist, found that a majority of patients answered questions incorrectly, suggesting that the RC of those PEMs was too difficult for the patients to read, comprehend, and apply the information.[30] It has been proven repeatedly in the past that PEMs are written at higher-than-recommended reading levels. Although researchers of these studies have hoped for change and improvement in PEM readability, our results highlight the persistent issue of inadequate readability of PEMs and emphasize the call for immediate action.

It is important to acknowledge the limitations of our study. Our evaluation software was not able to account for the inclusion of pictorials, graphs, and tables in the PEMs, which could be essential for patient comprehension. The exclusion of these visuals in our analysis may have artificially inflated the reading difficulty scores. We recommend that future studies use metrics that can assess the graphics and layout of PEMs to assess RC more comprehensively, such as with the Patient Education Materials Assessment Tool.[31] Moreover, our study was designed without patient participants and therefore evaluated RC indirectly. Although these metrics provided objective data, we recommend future studies to evaluate the patients' comprehension levels directly and compare readability measures to patient responses to increase the validity of the results. It is important to note that this is a limitation of other studies that use these measures to evaluate RC of PEMs, as the validity and reliability of these RC measures have not been well-studied for specifically health-related reading materials.[21] Another limitation is the generalizability of our findings, as our analysis was limited to English-language PEMs. Although there were ACR PEMs in Spanish, evaluating these PEMs was beyond the scope of this study, as linguistic differences in Spanish and English require different readability metrics to assess RC. Therefore, by only evaluating English PEMs, our study excludes a significant subset of patients who may experience low HL and poorer outcomes, particularly among different racial and ethnic groups. Access to patient-centric PEMs in patients' native languages is crucial for addressing health disparities. Additionally, our study focused specifically on PEMs from the ACR and JAMA websites because they are reputable

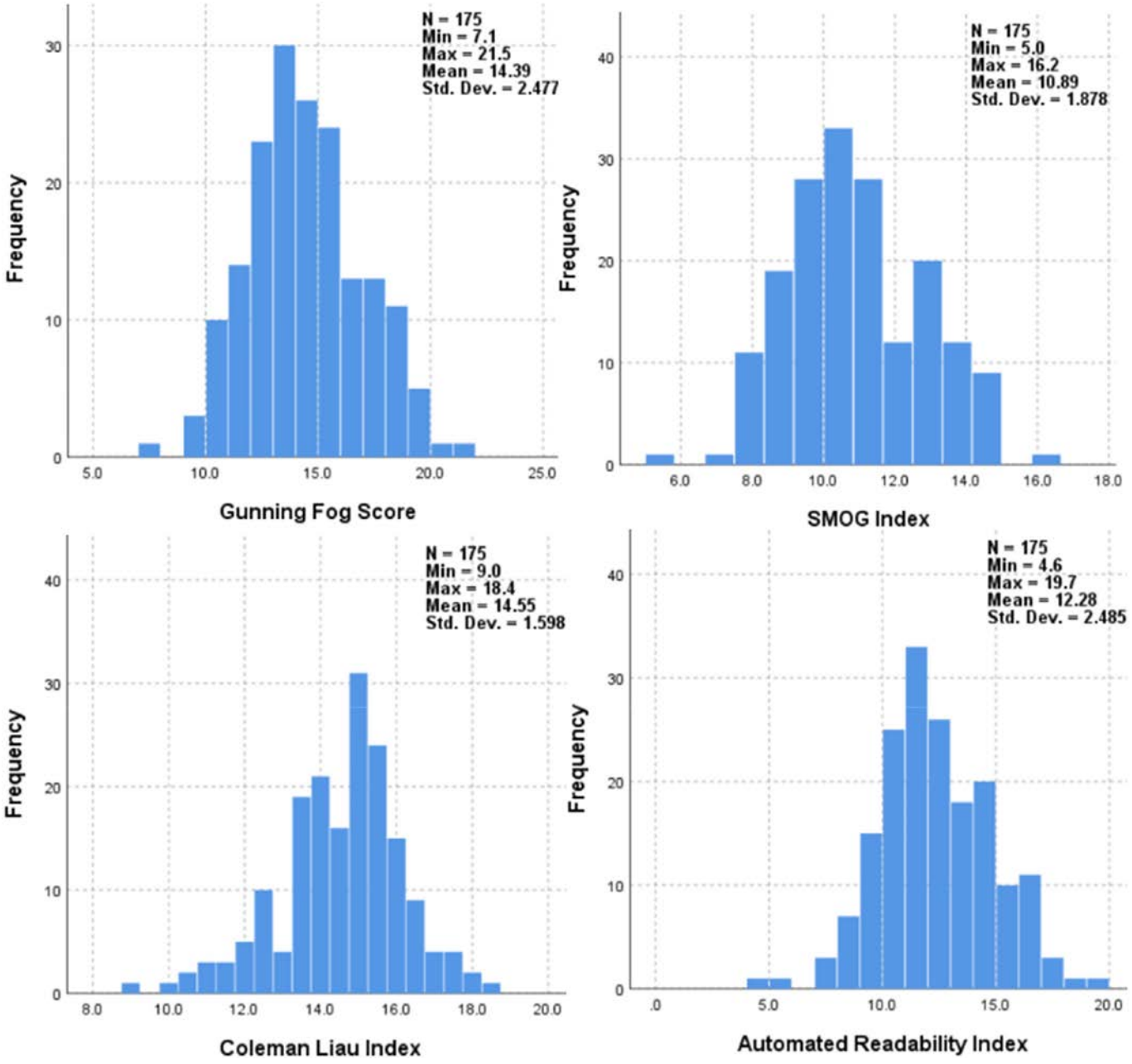


Figure 2. Distributions of all patient education materials for the Gunning Fog Score, SMOG, Coleman-Liau Index, and Automated Readability Index. SMOG, Simple Measure of Gobbledygook Index. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25473/abstract.

organizations that provide easy access to PEMs. However, we do not have data on how many patients or clinics actually use PEMs from these sources versus other sources. We understand that some patients may refer to disease-specific organizations, such as the Lupus Foundation of America or the American Heart Association; however, analyzing disease-specific organizations was beyond the scope of this study. We recommend future studies analyze the RC of PEMs available on these disease-specific organization websites. Additionally, we did not compare RC of PEMs based on disease complexity (such as lupus versus bursitis), as it was beyond the scope of our study to assign complexity to each disease process. However, we recommend a future study to analyze if difficulty in RC correlates with disease complexity. Additionally, we recognize that institutions may have their own built-in systems for PEMs or may access PEMs directly from their EMR system. However, we did not analyze these PEMs because they are not accessible to institutions that do not have the same EMR or open access to resources such as UpToDate. We recommend future studies analyze PEMs that are built into an institution's EMR. Lastly, our study was conducted in 2022, and although we have shown that there has been little progress in the readability of PEMs for rheumatology and internal medicine topics in the

past several years,[11,29,30] it is possible that if our study is repeated now, we could see some improvement in the RC of PEMs on these websites. We recommend that PEM RC continue to be studied, to advocate for organizations to make PEMs easier for patients to use them.

Despite these limitations, our study also possesses several strengths. We gathered a substantial number of PEMs from the ACR and JAMA websites, allowing for a comprehensive comparison of their readability metrics. We employed multiple standardized methodologies and metrics to assess readability, ensuring consistency in our findings. By using various metrics and considering the recommendations of national organizations such as the NIH and CDC, we aimed to provide a balanced and comprehensive assessment of the reading levels of PEMs, recommending a maximum reading level of seventh grade.

Overall, this study highlights that PEMs available on the ACR and JAMA websites are not patient-friendly, as they do not align with national organizations' recommendations for RC literacy levels. The RC gap poses challenges for patients with HL difficulties and may be detrimental to their health outcomes and health care use.

To address this issue, revising and improving the readability of existing PEMs on these websites should be prioritized. Efforts should be focused on aligning the materials with recommended readability metrics, to ensure they are accessible and understandable to patients with diverse literacy levels. We suggest going beyond just evaluating PEMs with the RC tools in this study and also directly involve patients from diverse HL backgrounds in the PEM revision and development process. Future development of PEMs must incorporate a patient-centered approach to create high-quality educational materials. Based on our results, we recommend the development of proper guidelines for the publication of new materials on the ACR and JAMA Network websites. We propose the development of a PEM-focused task force within each organization to ensure PEMs meet the appropriate RC criteria. For additional transparency, we suggest that each PEM on the ACR and JAMA websites contain a "readability ease score," ranging from "basic" to "advanced." For example, if PEMs are scored via FKGL, "basic" would correspond to grade levels below seventh grade, "fairly basic" would be assigned to seventh- and eighth-grade level, "fairly advanced" would be assigned to 9th- to 12-grade level, and "advanced" would be assigned to college level. We suggest against the inclusion of the actual grade level on PEMs to avoid confusing or stigmatizing patients. Rather, this "readability ease score" can help providers choose appropriate PEMs based on their patients' reading and education levels. We also understand that some patients may seek out more detailed and advanced information; therefore, transparent readability scores can ensure that patients can access PEMs that fit into their unique HL level. In addition to increasing transparency about the difficulty level of PEMS, this recommendation can also help create accountability for improving the PEMs' readability.

For the actual development of easily readable and understandable PEMs, the CDC's "Simply put" guide provides valuable information.[17] This resource emphasizes the importance of conducting a needs analysis, setting learning objectives, composing appropriate content, and evaluating the effectiveness of the materials. The guide contains specific recommendations for improving the RC of PEMs; for example, avoiding medical jargon by using "high blood pressure" rather than "hypertension." Attention should also be given to the language used, ensuring its suitability for the target patient group.[32] Again, we emphasize the direct involvement of patients in the revision of current PEMs and the development of new PEMs.

In conclusion, current PEMs from ACR and JAMA exceed the nationally recommended reading and comprehension levels of sixth to eighth grade. Our study demonstrates that PEMs from ACR and JAMA are at the high school or college level. Additionally, JAMA PEMs are more difficult for patients to read and comprehend compared to ACR PEMs. As effective PEMs help educate patients about their health, promote shared decision-making, increase medication compliance, lower health care costs, and improve health outcomes, it is essential that they are more patient-friendly. Therefore, current PEMs should be revised to the appropriate reading level, and new PEMs should be created in line with national recommendations.

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 Rustomji 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. Office of the Surgeon General (US); Office of Disease Prevention and Health Promotion (US). Panel 1: health literacy, literacy, and health outcomes. In: Proceedings of the Surgeon General's Workshop on Improving Health Literacy: September 7, 2006, National Institutes of Health, Bethesda, MD. Office of the Surgeon General (US); 2006.
2. Sørensen K, Van den Broucke S, Fullam J, et al; (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12(1):80.
3. Kutner M, Greenberg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy. Institute of Education Sciences website. 2006. Accessed May 1, 2022. http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483
4. Chen X, Hay JL, Waters EA, et al. Health literacy and use and trust in health information. J Health Commun 2018;23(8):724–734.
5. Scalia P, Durand MA, Elwyn G. Shared decision-making interventions: an overview and a meta-analysis of their impact on vaccine uptake. J Intern Med 2022;291(4):408–425.

6. Bartlett EE. How can patient education contribute to improved health care under prospective pricing? Health Policy 1986;6(3):283–294.
7. FastStats Arthritis. Centers for Disease Control and Prevention website. Updated 2024. Accessed December 26, 2024. https://www.cdc.gov/nchs/fastats/arthritis.htm.
8. Chronic Disease Indicators: Arthritis. Centers for Disease Control and Prevention website. 2024. Accessed December 26, 2024. https://www.cdc.gov/cdi/indicator-definitions/arthritis.html
9. Geryk LL, Blalock S, DeVellis RF, et al. Associations between patient characteristics and the amount of arthritis medication information patients receive. J Health Commun 2016;21(10):1122–1130.
10. Stossel LM, Segar N, Gliatto P, et al. Readability of patient education materials available at the point of care. J Gen Intern Med 2012;27(9):1165–1170.
11. Rhee RL, Von Feldt JM, Schumacher HR, et al. Readability and suitability assessment of patient education materials in rheumatic diseases. Arthritis Care Res (Hoboken) 2013;65(10):1702–1706.
12. Wasir AS, Volgman AS, Jolly M. Assessing readability and comprehension of web-based patient education materials by American Heart Association (AHA) and CardioSmart online platform by American College of Cardiology (ACC): how useful are these websites for patient understanding? Am Heart J Plus 2023;32:100308.
13. Kenner MM, Taylor ML, Dunn PC, et al. Primary care providers need a variety of nutrition and wellness patient education materials. J Am Diet Assoc 1999;99(4):462–466.
14. America's Health Literacy: Why We Need Accessible Health Information. US Department of Health and Human Services. 2008. Accessed January 10, 2024. https://www.ahrq.gov/sites/default/files/wysiwyg/health-literacy/dhhs-2008-issue-brief.pdf
15. Weiss BD. Health Literacy: A Manual For Clinicians. American Medical Association; 2003. Accessed January 10, 2024. http://lib.ncfh.org/pdfs/6617.pdf
16. Clear Communication: Clear and Simple National Library of Medicine Website. Accessed December 26, 2024. https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication/clear-simple.
17. Centers for Disease Control and Prevention. Simply put: A guide for creating easy-to-understand materials. Centers for Disease Control and Prevention website. Accessed January 10, 2024. https://stacks.cdc.gov/view/cdc/11938.
18. Han A, Carayannopoulos AG. Readability of patient education materials in physical medicine and rehabilitation (PM&R): a comparative cross-sectional study. PM R 2020;12(4):368–373.
19. Nattam A, Vithala T, Wu TC, et al. Assessing the readability of online patient education materials in obstetrics and gynecology using traditional measures: comparative analysis and limitations. J Med Internet Res 2023;25:e46346.
20. Wang LW, Miller MJ, Schmitt MR, et al. Assessing readability formula differences with written health information materials: application, results, and recommendations. Res Social Adm Pharm 2013;9(5):503–516.
21. Carter B, Nayak K, Vembenil I. The accuracy of readability formulas in health content: a systematic review. J Healthc Commun 2024;9(1):9002.
22. Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155(2):97–107.
23. Bennett IM, Chen J, Soroui JS, et al. The contribution of health literacy to disparities in self-rated health status and preventive health behaviors in older adults. Ann Fam Med 2009;7(3):204–211.
24. Fleary SA, Ettienne R. Social disparities in health literacy in the United States. Health Lit Res Pract 2019;3(1):e47–e52.
25. Coughlin SS, Vernon M, Hatzigeorgiou C, et al. Health literacy, social determinants of health, and disease prevention and control. J Environ Health Sci 2020;6(1):3061.
26. Sheridan SL, Halpern DJ, Viera AJ, et al. Interventions for individuals with low health literacy: a systematic review. J Health Commun 2011;16 suppl 3:30–54.
27. Dore A, Dye J, Hourani L, et al. Incorporating the health literacy universal precautions toolkit quick start in academic rheumatology practices: Carolina Fellows Collaborative. 2013 American College of Rheumatology/Association of Rheumatology Health Professionals Annual Meeting; October 25–30, 2013; San Diego, CA.
28. Hirsh J, Wood P, Keniston A, et al. Universal health literacy precautions are associated with a significant increase in medication adherence in vulnerable rheumatology patients. ACR Open Rheumatol 2020;2(2):110–118.
29. Rooney MK, Santiago G, Perni S, et al. Readability of patient education materials from high-impact medical journals: a 20-year analysis. J Patient Exp 2021;8:2374373521998847.
30. Oliffe M, Thompson E, Johnston J, et al. Assessing the readability and patient comprehension of rheumatology medicine information sheets: a cross-sectional Health Literacy Study. BMJ Open 2019;9(2):e024582.
31. Shoemaker SJ, Wolf MS, Brach C. Development of the Patient Education Materials Assessment Tool (PEMAT): a new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Couns 2014;96(3):395–403.
32. Whittington R. Top tips for creating patient education material. Rx Communications. 2012. Accessed May 14, 2024. https://www.rxcomms.com/blog/effective-patient-education-materials

Arthritis Care & Research
Vol. 77, No. 5, May 2025, pp 685–687

American College of Rheumatology
Empowering Rheumatology Professionals

LETTER

DOI 10.1002/acr.25474

Integrating patient advocacy groups in the development of clinical practice guidelines: comment on the article by Johnson et al

To the Editor:

The Sjögren's Foundation applauds the American College of Rheumatology (ACR) and the American College of Chest Physicians (CHEST) for their multispecialty collaboration in developing clinical practice guidelines (CPGs) for autoimmune disease–related interstitial lung diseases.[1] We would like to note that the Sjögren's Foundation pulmonary guidelines[2] expand on one of the autoimmune diseases included in this publication, Sjögren disease. Our guidelines, which complement the ACR/CHEST guidelines, involved multispecialty collaboration across pulmonology, rheumatology, and oncology as well as patient input at every stage.

CPGs play a vital role in standardizing care by ensuring a unified set of minimum criteria for patient management and highlight needed research. Such guidelines also aid in securing appropriate diagnostic and treatment modalities from insurance providers, making them a critical element in improving patient outcomes. As widely referenced and used tools by clinicians, CPGs have the potential for significant impact on patient health. Given their importance, it is essential to provide the most comprehensive expert guidance possible to inform day-to-day clinical decision-making. Exploring alternative methodologies that maintain high standards while reducing the burden on contributors could accelerate the development and release of new CPGs.[3,4]

Patient advocacy organizations (PAGs) serve as trusted sources of information within their respective communities and often have content experts on their leadership teams. We commend professional societies, such as ACR and CHEST, for leading efforts to create CPGs. However, it is unlikely that these societies alone will be able to address every management dilemma across all conditions. By partnering with PAGs, gaps in CPG development can be filled more efficiently, ensuring both timely creation and regular updates as new evidence emerges. Disease-specific organizations bring added value that further patient and provider disease-centric expertise. The involvement of PAGs in the CPG development process, or at least endorsement of the CPGs by the respective PAGs, might also help increase clinical uptake of the guidelines. We support increased PAG consideration of incorporating the creation and updating of CPGs into their organizations' missions and call on professional societies to actively engage PAGs throughout the CPG development process.

The authors below are members of the Topic Review Group for the Sjögren's Foundation Clinical Practice Guidelines, Pulmonary Manifestations of Sjögren's.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25474.

Nishant Gupta, MD
University of Cincinnati and Veterans Affairs Medical Center
Cincinnati, Ohio
Steven E. Carsons, MD
New York University Winthrop Hospital and New York University Langone Health
New York, New York
Nancy L. Carteron, MD
University of California, Berkeley and University of California, San Francisco
Robert Hal Scofield, MD
Oklahoma Medical Research Foundation and University of Oklahoma Health Sciences Center and Oklahoma City Department of Veterans Affairs Medical Center
Oklahoma City
Augustine S. Lee, MD
Mayo Clinic
Jacksonville, Florida
Donald E. Thomas, MD
Uniformed Services University of the Health Sciences
Bethesda, Maryland and Arthritis and Pain Associates of Prince Georges County Prince Georges County
Greenbelt, Maryland
Teng Moua, MD
Mayo Clinic
Rochester, Minnesota
Kamonpun Ussavarungsi, MD
Cleveland Clinic
Cleveland, Ohio
E. William St Clair, MD
Duke University Medical Center
Durham, North Carolina
Richard Meehan, MD
National Jewish Health
Denver, Colorado
Kieron Dunleavy, MD
MedStar Georgetown University Hospital
Washington, DC
Matt Makara, MPH
Katherine Morland Hammitt, MA
khammitt@sjogrens.org
Sjögren's Foundation
Reston, Virginia

1. 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 Care Res (Hoboken) 2024;76(8):1051–1069.

2. Lee AS, Scofield RH, Hammitt KM, et al; Consensus Expert Panel (CEP) Members. Consensus guidelines for evaluation and management of pulmonary disease in Sjögren's. Chest 2021;159(2): 683–698.
3. Wilson KC, Schoenberg NC, Raghu G. Idiopathic pulmonary fibrosis guideline recommendations. Need for adherence to institute of medicine methodology? Ann Am Thorac Soc 2019;16(6):681–686.
4. Wilson KC, Schoenberg NC, Cohn DL, et al. Community-acquired pneumonia guideline recommendations-impact of a consensus-based process versus systematic reviews. Clin Infect Dis 2021;73(7):e1467–e1475.

DOI 10.1002/acr.25475

Reply

To the Editor:

We thank the authors for their important letter to the editor in response to the publication of the American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) guidelines for the screening,[1,2] monitoring,[1,2] and treatment[3,4] of interstitial lung disease (ILD) in people with systemic autoimmune rheumatic diseases. The ACR agrees with the authors' stated request and applauds all guideline developers' efforts to include as many interested parties in guideline development as possible, to better reflect and address broad perspectives.

The ACR places a high priority on developing methodologically rigorous, evidence-based clinical practice guidelines that take into consideration multiple viewpoints and varied expertise and experiences, including those of patients, in a systematic and transparent manner.[5] From the outset of ACR guideline development, the inclusion of people with the diseases that are the topic of the guideline (and parents or caregivers, if the guideline topic is pediatric) is prioritized. Patients (including parents or caregivers, if a pediatric guideline) comprise each ACR guideline's patient panel, which discusses patient values and preferences related to outcomes, evidence, and drafted recommendation statements.[5–7] Some of these patient panelists also sit on the voting panel to represent the views of the patient panel in voting discussions and decisions, which take into consideration tradeoffs between the benefits and harms of alternative management strategies.[5–7] Patients may self-nominate, be identified from the clinics of physicians not serving on the guideline development team (to help ensure patients feel open to speak freely by avoiding conflicts of interest with those serving on the guideline development team), or be identified through patient advocacy groups. Indeed, the ACR has collaborated with patient advocacy groups for several guidelines across a range of rheumatic diseases.[8–10] The input of patients is particularly important in the setting of guidelines informed by low certainty evidence, as these recommendations are particularly sensitive to patients' values and preferences. In the ACR/CHEST ILD guidelines, patients expressed the desire to undergo screening and monitoring for the early detection of ILD and highlighted the importance of good communication, particularly related to adverse effects of medications and goals of therapy.[11]

The ILD guideline project was supported by the American College of Rheumatology.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25475.

Sindhu R. Johnson, MD, PhD
Sindhu.Johnson@uhn.ca
University of Toronto and Toronto Western Hospital and Mount Sinai Hospital
Toronto, Ontario, Canada
Amy Turner
American College of Rheumatology
Atlanta, Georgia
Elana J. Bernstein, MD, MSc
Columbia University Irving Medical Center
New York, New York

1. Johnson SR, Bernstein EJ, Bolster MB, et al. 2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) guideline for the screening and monitoring of interstitial lung disease in people with systemic autoimmune rheumatic diseases. Arthritis Care Res (Hoboken) 2024;76(8): 1070–1082.
2. Johnson SR, Bernstein EJ, Bolster MB, et al. 2023 American College of Rheumatology (ACR)/American College of Chest Physicians (CHEST) guideline for the screening and monitoring of interstitial lung disease in people with systemic autoimmune rheumatic diseases. Arthritis Rheumatol 2024;76(8):1201–1213.
3. 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 Care Res (Hoboken) 2024;76(8):1051–1069.
4. 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.
5. Johnson SR, Turner AS, Goodman SM. How the American College of Rheumatology develops guidelines. Rheum Dis Clin North Am 2022; 48(3):579–588.
6. Fraenkel L, Miller AS, Clayton K, et al. When patients write the guidelines: patient panel recommendations for the treatment of rheumatoid arthritis. Arthritis Care Res (Hoboken) 2016;68(1): 26–35.
7. Goodman SM, Miller AS, Turgunbaev M, et al. Clinical practice guidelines: incorporating input from a patient panel. Arthritis Care Res (Hoboken) 2017;69(8):1125–1130.
8. Singh JA, Guyatt G, Ogdie A, et al. Special Article: 2018 American College of Rheumatology/National Psoriasis Foundation guideline for the treatment of psoriatic arthritis. Arthritis Care Res (Hoboken) 2019;71(1):2–29.

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.
10. Beukelman T, Patkar NM, Saag KG, et al. 2011 American College of Rheumatology recommendations for the treatment of juvenile idiopathic arthritis: initiation and safety monitoring of therapeutic agents for the treatment of arthritis and systemic features. Arthritis Care Res (Hoboken) 2011;63(4):465–482.
11. Mirza RD, Bolster MB, Johnson SR, et al. Assessing patient values and preferences to inform the 2023 American College of Rheumatology/American College of Chest Physicians interstitial lung disease guidelines. Arthritis Care Res (Hoboken) 2024;76(8):1083–1089.