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. 10 October 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 • OCTOBER 2025 • NO. 10

Cover image: The image on the cover (Slabbynck et al; pages 1184–1193) shows possible mechanisms of NLRP3 inflammasome activation by cholesterol crystals in macrophages and neutrophils in the alveolar space and their cross talk. Image made with BioRender, adapted from Tall et al, CC BY 4.0; ref. 51.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1167–1174
DOI 10.1002/acr.25537

American College of Rheumatology
Empowering Rheumatology Professionals

REVIEW ARTICLE

Metabolic Consequences of Rheumatoid Arthritis

Stevie Barry,[1] Emily Sheng,[1] and Joshua F. Baker[2]

Patients with rheumatoid arthritis (RA) may have metabolic disruption, which can contribute to adverse long-term outcomes, for multiple reasons. Patients with RA appear to have a higher risk of sarcopenia, type 1 and type 2 diabetes mellitus, metabolic syndrome, and hypertension. Systemic inflammation in RA can cause a "lipid paradox," with reduced low-density lipoprotein being associated with higher rates of cardiovascular disease. In this review, we discuss changes to body composition, insulin resistance, lipids, and blood pressure that often occur in patients with RA. We examine the current understanding of the mechanisms underlying disruptions in metabolic pathways in RA, their clinical effects, and how treatment affects these changes.

Introduction

The mechanisms involved in how organisms process, store, and use energy are among the most well-regulated and important. In times of disease, disruptions to metabolism are observed because of a need to shift resources to areas of the body in greatest need. In the modern era, energy overabundance is the most common source of metabolic disturbance due to excess adiposity and related adipose tissue inflammation, which can result in the disruption of normal metabolic pathways. Patients with rheumatoid arthritis (RA) may have metabolic disruption, which can contribute to adverse long-term outcomes, for multiple reasons. These include those related to energy overabundance and excess adiposity, as well as those related to inflammation and high energy use, resulting in a complex picture (Figure 1). In this review, we aim to review the ways that systemic inflammation and other manifestations and complications of RA can lead to metabolic disturbances and the clinical implications of such changes for clinicians.

Changes to body composition

Body mass index (BMI) is the most feasible and widely used method to assess metabolic health in clinical practice. BMI is correlated with excess adiposity and the metabolic complications of excess adiposity, though not perfectly. An underweight BMI can also be a sign of undernutrition or cachexia. Patients with RA are generally observed to have higher BMI compared to the general population, particularly in the recent era,[1] though they may have modestly lower BMI compared to those with arthritic conditions, such as osteoarthritis (OA) and psoriatic arthritis.[2–4]

Changes in weight over time are also informative. It is normal for adults to gain weight through middle age and lose weight later in life.[5] Changes in weight over the lifespan may be due to effects of behavior and aging or may be related to the development of chronic illnesses (eg, diabetes mellitus [DM], heart failure). Weight cycling, or the frequent alternating between periods of weight gain and weight loss, has also been observed as a predictor of adverse outcomes.[6–8] For example, weight cycling among patients with RA is associated with higher rates of cardiovascular events.[6] These observations may be the result of direct metabolic and proinflammatory effects of weight fluctuation, as has been suggested in some mouse studies,[9,10] or may be related to changes in health status and the accumulation of comorbidity.

An individual's weight is composed of both lean tissue (organs, muscle) and adipose tissue. These individual compartments can be measured separately through multiple methods. These techniques have allowed investigators to demonstrate that the effects of muscle and fat mass on health outcomes are often distinct or even opposing. Prior studies have demonstrated higher fat mass in patients with RA, including higher levels of visceral fat, on average.[11,12] However, severe RA is often observed to be associated with a significant loss of visceral fat, presumably due to cachexia.[11,13] Our group recently used a new approach to defining body composition phenotypes in RA.[14,15] In this study, the rates of obesity in RA were similar to those

Supported by U.S. Department of Veterans Affairs Merit Award Rehabilitation Development and Translation Merit Award (awards I01 RX003644 and I01 RX004770 to Dr Baker) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant R01-AR-083584 to Dr Baker).

[1]Stevie Barry, MB, BCh, BAO, Emily Sheng, BA: Perelman School of Medicine, University of Pennsylvania, Philadelphia; [2]Joshua F. Baker, MD, MSCE: Perelman School of Medicine, University of Pennsylvania and Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania.

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25537.

Address correspondence via email to Joshua F. Baker, MD, MSCE, at Joshua.baker@pennmedicine.upenn.edu.

Submitted for publication February 28, 2025; accepted in revised form March 19, 2025.

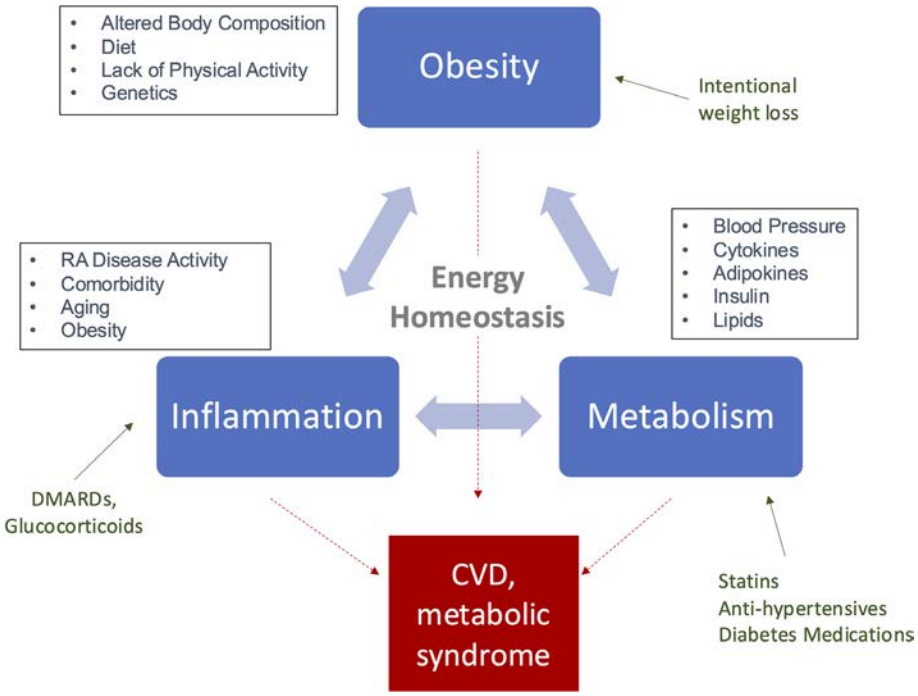


Figure 1. Bidirectional relationship between obesity, inflammation, and metabolism, which collectively contribute to metabolic syndrome and CVD in RA. Listed in dark green are potentially beneficial interventions for each component. Long-term adverse outcomes relevant to metabolic disruption are shown in the red box. CVD, cardiovascular disease; DMARD, disease-modifying antirheumatic drug; RA, rheumatoid arthritis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25537/abstract.

in reference populations, though there was a higher rate of sarcopenia and sarcopenic obesity.[16] Deficits in muscle mass compared to controls tend to occur among patients with RA with lower adiposity, suggesting that they develop during periods of cachexia.[17] Infiltration of fat into muscles is also observed in RA and is also more prominent compared to controls among those with low adiposity.[17–19] Although not well studied, the loss of muscle mass and muscle quality early in the disease during periods of cachexia may contribute to adverse long-term metabolic outcomes over the lifespan by affecting the health of muscle tissue, a major regulator of metabolism.

Although rates of obesity appear to be similar or somewhat greater among patients with RA compared to the general population, it is clear that the inflammatory disease, reductions in physical activity, and perhaps medication toxicities can result in measurable and important changes to body composition, including muscle loss, that may impact metabolic outcomes over the longer term. Because metabolic changes related to disease processes may occur, as outlined in the next section, weight and even excess adiposity may not clearly correlate with metabolic dysfunction and adverse outcomes in RA compared to what is observed in the general population.[20]

Chronic systemic inflammation and insulin resistance

DM is a highly prevalent chronic condition that is widely believed to stem from obesity-related chronic low-grade systemic inflammation leading to insulin resistance (IR). Obesity is characterized by the infiltration of immune cells, dominated by proinflammatory M1-like polarized macrophages, into adipose tissue and other major metabolic tissues.[21] These immune cells secrete proinflammatory cytokines, leading to IR through a variety of signaling cascades affecting insulin receptors, lipid metabolites, and reactive oxygen species (ROS).[21] This association between inflammation and IR may stem from the evolutionary need to mobilize energy in the context of infection or injury.

Because chronic tissue inflammation has become an increasingly well-understood cause of IR broadly,[21] studies have focused on evaluating whether patients with active RA exhibit IR compared to controls.[22] A cross-sectional study found that higher RA disease activity, assessed with the Disease Activity Score in 28 joints, corresponded to greater IR as measured by the Homeostasis Model Assessment of Insulin Resistance (HOMA-IR).[23] Another study, however, found no difference in IR in patients with untreated early RA compared to healthy controls despite observing a difference in measures of systemic inflammation.[24]

The underlying mechanism through which systemic inflammation leads to IR is likely at least partly related to the direct effect of adipokines and cytokines on glucose metabolism. Research using collagen-induced arthritis mouse models found that RA-induced adipose tissue inflammation led to the down-regulation of insulin-dependent glucose transporter 4 (GLUT-4) and insulin receptor substrates.[25] Synovial biopsies from patients with RA

and OA also identified significantly increased GLUT-1 expression and decreased GLUT-4 expression in the RA samples.[26] RA-related elevations in tumor necrosis factor (TNF), interleukin-1β (IL-1β), and IL-6 cytokine levels also interfere with insulin receptor signaling.[22,27] Notably, TNF promotes serine phosphorylation of insulin receptor substrate 1, effectively inhibiting glucose uptake,[28] and IL-1β activates the NF-κB pathway to interfere with insulin receptor signaling.[22] IL-6 is shown to have a dual role as both an anti-inflammatory and a proinflammatory effect depending on the tissue, with prolonged IL-6 elevations reducing insulin sensitivity in adipocytes.[22]

Epidemiologic studies have shown that RA is associated with a higher risk of IR conditions, including DM and metabolic syndrome.[29–31] A meta-analysis pooling data from multiple case-control studies found RA was associated with a higher risk of both type 1 and type 2 DM.[32] A case-control study found a significantly higher prevalence of metabolic syndrome, a risk that was associated with increasing disease duration.[33] However, a recent study conducted in administrative data found that patients with RA had a lower rate of incident type 2 DM than patients with hypertension (HTN), OA, or psoriatic arthritis.[34] Differences between studies may stem from differences in classification of disease, comparator groups, modeling approaches, and perhaps the severity of RA disease.

Higher RA disease activity and systemic inflammation have also been independently associated with a higher risk of incident DM among patients with RA independent of traditional risk factors such as age, sex, and BMI.[35] For example, higher levels of C-reactive protein (CRP), IL-1, IL-4, and IL-6 were each found to be significantly associated with the risk of DM, supporting the hypothesis that systemic inflammation related to RA is a risk factor for DM and metabolic syndrome. These observations raise the obvious question as to whether therapeutic targets aimed at reducing RA inflammation may reduce the risk of DM and improve the control of DM among patients with both conditions.[36]

The increased adoption and use of biologic disease-modifying antirheumatic drugs (bDMARDs) in RA therapy in recent years has provided an opportunity to evaluate the effect of immunologic pathways on IR and related outcomes in patients with RA. TNF inhibitors (TNFi), for example, have been associated with a lower risk of DM onset than methotrexate and other non-bDMARDs (eg, sulfasalazine, leflunomide, cyclosporine)[37] and with improvements in HOMA-IR and the Quantitative Insulin Sensitivity Check.[38]

Additional recent work found that anakinra, an IL-1 receptor antagonist, improved glycemic control among patients with DM. The TRACK clinical trial demonstrated that patients with RA and type 2 DM randomized to receive anakinra had a significant reduction in hemoglobin A1c (HbA1c) levels at both three and six months compared to those randomized to TNFi.[39] Long-term follow-up of this study indicated that benefits persisted, with a notable reduction in the need for antidiabetic medications in the anakinra group compared with the TNFi group.[40] Interestingly, treatment with IL-1 therapy (canakinumab) in a non-RA cohort of patients at risk of cardiovascular disease showed no reduction in the risk of DM, though there was a short-term reduction in HbA1c levels.[41] Sarilumab, an IL-6 inhibitor, has also been associated with greater reductions in HbA1c levels compared to TNFi or conventional DMARDs in patients with both RA and DM in post hoc analyses of clinical trial data.[42]

It is interesting to consider that there may be a bidirectional relationship between IR, metabolic dysfunction, and RA disease activity, as well as treatment response. For example, obesity and metabolic syndrome are consistently associated with reduced response rates to advanced therapies.[43,44] However, whether refractory inflammation explains this relationship remains controversial.

Thus, there is evidence that treatment of active RA with bDMARDs (TNFi and IL-1 and IL-6 inhibitors) may offer benefits for glycemic control. However, more research is needed to identify management strategies for RA that consider glycemic control among those with both conditions. It is attractive to consider IR along with other metabolic outcomes as a target for treating RA; however, more evidence to support benefits is needed.

Inflammation and changes in circulating lipids

There is a complex interplay between lipids, inflammation, and cardiovascular disease in RA. Systemic inflammation, particularly in RA, is increasingly recognized as a potent driver of atherosclerosis, and the magnitude of increased risk in patients with RA has been estimated to be equivalent to that of DM.[45] Recent trials, such as the Canakinumab Anti-inflammatory Thrombosis Outcome Study (CANTOS), have demonstrated fewer recurrent cardiovascular events in patients treated with an inhibitor of IL-1β, suggesting a causal effect of inflammation on cardiovascular risk.[46]

Hyperlipidemia is a well-recognized risk factor for cardiovascular disease; however, inflammation itself directly affects lipid processing, resulting in lower levels of low-density lipoprotein (LDL) and total cholesterol (TC). This observation is not unique to RA and has been demonstrated in many inflammatory states, including sepsis and endotoxin exposure, and is observed in the postsurgical state.[47] Low levels of lipoproteins have not only been demonstrated in patients with RA but have also been observed in preclinical RA, as long as five years before diagnosis.[48] The effect of inflammation on both cardiovascular risk and lipid processing may help to explain the "lipid paradox" that has been observed in RA. The "paradox" refers to the observation that low LDL levels are often associated with higher rates of cardiovascular disease in patients with the disease.[48] Recently, very low LDL levels in patients with RA were found to be associated with significantly worse atherosclerotic burden on computed tomography

coronary angiography compared to a normal LDL level (above 70 mg/dL).[49]

Recent work has aimed to identify the mechanisms and pathways that lead to low lipid levels and how activation of these pathways may lead to higher cardiovascular risk. Potential mechanisms for the reductions in high-density lipoprotein (HDL) and LDL levels observed in the setting of inflammation include an increased rate of metabolism of cholesterol in HDL and LDL, suppressed production, and accelerated deposition of LDL as oxidized LDL in arterial walls.[49–51] One study found that antibodies to oxidized LDL were associated with coronary plaque burden, suggesting that excessive oxidization of LDL may also occur and could contribute to long-term risk.[49] Other changes in lipoprotein structure and function may also occur. The protein "cargo" associated with HDL in active RA has been shown to be altered toward proinflammatory proteins, which may negate or reverse HDL's typical anti-inflammatory effect on atherosclerotic plaques.[52] Recent evidence has suggested that reduced activity of one HDL-associated enzyme, paraoxonase 1 (PON-1), is associated with a higher risk of cardiovascular events in patients with RA.[53] Further, greater PON-1 activity was associated with less joint inflammation in mouse models.[54]

The reversal of inflammation can also affect circulating lipoproteins. Elevations in TC, LDL, and HDL levels (often without a change in the TC/HDL ratio) have been demonstrated in patients with RA when treated with DMARDs, including methotrexate and sulfasalazine, and multiple biologics and targeted synthetic therapies, including multiple TNFi, tocilizumab, and tofacitinib.[47] Treatment-related reductions in CRP levels have been correlated with increases in HDL and TC levels and normalization of cholesterol metabolism.[55] Methotrexate, multiple TNFi, tocilizumab, and tofacitinib have all demonstrated changes in markers associated with a shift from proinflammatory to anti-inflammatory HDL.[47]

Screening of lipids is largely performed to help predict cardiovascular risk, a practice that is made more difficult in patients with systemic inflammation. RA disease activity has itself long been known to correlate with a higher risk of cardiovascular events and cardiovascular mortality.[56] In addition, treatment of inflammatory disease with DMARD therapy can reduce cardiovascular risk in RA, though studies to date have largely been observational in nature. A Swedish study demonstrated a decreasing cardiovascular disease risk with use of any DMARD during the first year of diagnosis.[57] In another study, EULAR good responders to TNFi, but not moderate responders or nonresponders, had the same risk of cardiovascular events as the general population.[58] EULAR guidelines based on observational data suggested that TNFi therapies might have the greatest reduction in cardiovascular risk.[59] However, in a recent randomized trial, TNFi were not superior to conventional DMARDs in reducing vascular inflammation on positron emission tomography, which may raise concerns of bias in prior observational studies.[60] A meta-analysis of observational studies concluded that tocilizumab and abatacept have a similar risk reduction compared to TNFi.[61] Methotrexate in observational studies appears to outperform other conventional synthetic DMARDs, though this may suffer from confounding by indication.[59] Overall, it appears that disease control is critical, but there is limited evidence to support one DMARD over another.

In response to a very notable rise in lipid levels in phase 3 trials of tocilizumab and tofacitinib, noninferiority trials were conducted.[62] Tocilizumab was shown to be noninferior to TNFi regarding cardiovascular events; however, tofacitinib versus TNFi was shown to cross the upper end of the major adverse cardiovascular events risk confidence interval in the ORAL Surveillance trial.[63,64] Although the failure to meet criteria for noninferiority has raised some concerns, it remains unclear whether JAK inhibitor (JAKi) use raises the risk of cardiovascular events or whether TNFi may simply outperform tofacitinib regarding cardiovascular risk reduction.[65] Although it is clear from the body of evidence that reducing inflammation appears to reduce cardiovascular risk, the independent effect of concurrent increases in lipid levels remain unclear. A large randomized trial suggested a benefit of statin use in patients with RA, with greater reductions in LDL levels associated with greater benefit. Interestingly, reductions in CRP levels with statin use suggest a potential benefit on inflammatory pathways as well.[66] Further study is needed to help clinicians determine when to initiate lipid-lowering therapy for lipid abnormalities in RA.

Attempts have been made to develop risk calculators to identify patients with RA who would benefit from lipid-lowering therapy, including a recommendation by EULAR of multiplying standard scores by 1.5 for all patients with RA. However, no score has yet been widely implemented, and some evidence suggests newer models are not meaningfully superior to traditional risk models.[67] Assessing carotid intimal thickness has been suggested by EULAR because the presence of plaque on carotid ultrasound is an immediate statin indication.[59] A recent study also suggested that coronary calcium scores can help identify patients with evidence of atherosclerosis who have low LDL levels.[68] Although RA-specific risk prediction tools have not proven highly useful, it seems rational to consider those with high disease activity and severity as particularly high risk, even when their risk is low by traditional tools. Importantly, many patients with RA at relatively high cardiovascular risk may not reach treatment thresholds for statin use as a result of the lipid paradox. Thus, lower treatment targets may be appropriate to consider.

Inflammation and blood pressure

HTN is a frequently encountered comorbidity in RA. The prevalence in individuals with RA appears to exceed that in the general population and may be underdiagnosed and undertreated.[69,70] Although higher rates of obesity and limitations of

physical activity likely contribute to this increased prevalence, there is also increasing evidence of a direct link between systemic inflammation and HTN. In non-RA studies, CRP has been shown to correlate with HTN.[71] HTN is also more prevalent in individuals with other inflammatory and autoimmune conditions, such as Sjögren disease[72] and systemic lupus erythematosus,[73] compared to the general population. Mice studies have demonstrated that suppression of both the innate and adaptive branches of the immune system can prevent the induction of HTN.[74]

The role of inflammation in the development of HTN is far from fully elucidated, but unifying theories have been proposed. One such theory suggests that vascular tissue injury may result in an inflammatory cascade, perhaps by promoting the development of ROS. ROS have been shown to cause kidney damage induced by angiotensin II release, increase hypothalamic vasopressin release, and result in endothelial dysfunction that increases arterial stiffness and thus peripheral resistance.[74,75] Cytokines, immune cells, and other inflammatory mediators may also directly interact with vascular walls and cause endothelial dysfunction.[74]

It is important to note that, although basic science and human data suggest an association between inflammation and HTN, there are no human trials supporting intervention on inflammatory pathways. Recent data from the CANTOS trial suggested that blocking IL-1 had no benefit regarding blood pressure (BP).[76] If inflammation does represent a causal mechanism for the development of HTN, it remains uncertain whether and how systemic inflammation in patients with RA might activate and/or perpetuate these pathways and whether intervention may be beneficial.

Systemic inflammation, even in the absence of evidence of active joint disease, is associated with HTN in RA. One study in patients with RA and matched controls demonstrated a higher prevalence of HTN in patients with RA using 24-hour ambulatory BP. This study also noted that, even in patients with low clinical disease activity, there was a correlation between systemic inflammation (as measured by a panel of approximately 70 inflammatory mediators) and increased 24-hour ambulatory BP. Confounding related to obesity, visceral adiposity, and other comorbidities may have contributed to this result because these factors were not considered in these analyses.[77] Other studies have suggested that inflammatory-mediated endothelial dysfunction, as measured by impaired vasodilatory response to nitric oxide, may serve as a link between inflammation and HTN in RA.[78] Endothelial dysfunction and the resulting increased arterial stiffness were noted in patients with RA when compared to non-RA controls and appeared to be reversed with DMARD therapy.[79,80]

Few studies have directly evaluated the effects of treatment of RA with DMARDs on BP and rates of HTN. We studied patients with RA before versus after DMARD initiation in a large US Department of Veterans Affairs database and demonstrated a reduction in post-treatment BP. Even among those who did not receive antihypertensive therapy throughout the study period, there was a plateauing of and some reduction in BP.[81] These observational data are supported by a recent meta-analysis of randomized controlled trials and observational studies looking at risk factors for HTN in RA.[82]

It has been suggested that methotrexate increases adenosine production, which acts as a vasodilator.[83] Leflunomide appears to have a more modest reduction in BP and may increase BP in some cases. In our study, leflunomide was associated with less reduction in BP compared to methotrexate and was associated with a higher risk of incident HTN.[81] It has been suggested that leflunomide may increase BP by increasing sympathetic tone.[81] Of note, the most recent update to EULAR guidelines on cardiovascular risk in RA does not recommend for or against a particular DMARD therapy for patients with HTN given a lack of evidence.[59]

Although HTN is a known side effect of glucocorticoids, this must be considered in the context of the anti-inflammatory effect that might be expected to reduce BP in RA. Wolfe and Michaud studied a cohort of 17,738 patients with RA that demonstrated 49.1% of prednisone users had HTN compared to 45.6% of those who did not receive prednisone during the course of the seven-year study.[84] A UK retrospective study of patients with RA from 1992 to 2019 also demonstrated an increased incidence of HTN in those who had been prescribed the equivalent of ≥7.5 mg of prednisolone daily, suggesting a possible dose-dependent effect.[85] However, these studies suffer from

Table 1. Expected effects of RA inflammation, treatments, and comorbidities on individual clinical assessments of metabolic health in practice*

Metabolic changes	Active disease	Changes with RA treatment	Prednisone use	Aging	Physical inactivity
"Bad" cholesterol: LDL, TG	Low	High	High	High	High
"Good" cholesterol: HDL	Low	High	Low	Low	Low
Dysfunctional HDL	High	Low	Unknown	High	High
Blood pressure	High	Low	High	High	High
Insulin resistance	High	Low	High	High	High
Weight/visceral fat	Low	High	High	High to low	High

* For example, those with more severe and active RA inflammation may experience low lipoprotein levels, whereas effectively treated individuals who are physically inactive may have higher levels. HDL, high-density lipoprotein; LDL, low-density lipoprotein; RA, rheumatoid arthritis; TG, triglycerides.

confounding by indication. In our study, we observed an overall reduction in BP in those initiating prednisone.[81]

Inflammation is increasingly recognized as a potential driver of HTN. Although HTN is common in RA, the mechanism(s) behind this and how treatment may be best used to address HTN in RA are not yet clear.

Conclusions

There are myriad metabolic consequences of systemic inflammation that are highly relevant for patients with RA and other inflammatory conditions. A number of frequently encountered issues in RA and their effects on individual metabolic changes are outlined in Table 1, summarizing the complex interplay between RA and the many clinical assessments of metabolic health often made in clinical practice. These metabolic consequences can have a considerable morbidity and mortality burden and deserve attention. Although current evidence does not support metabolic targets for treatment, they represent an important potential off-target benefit that should be considered by clinicians. Further research into the metabolic changes in RA would likely improve our understanding of how best to manage patients with RA and also further our understanding of these common metabolic conditions in the general population.

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 Baker 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. Albrecht K, Richter A, Callhoff J, et al. Body mass index distribution in rheumatoid arthritis: a collaborative analysis from three large German rheumatoid arthritis databases. Arthritis Res Ther 2016;18(1):149.
2. Labitigan M, Bahče-Altuntas A, Kremer JM, et al. Higher rates and clustering of abnormal lipids, obesity, and diabetes mellitus in psoriatic arthritis compared with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2014;66(4):600–607.
3. Ogdie A, Harter L, Shin D, et al. The risk of fracture among patients with psoriatic arthritis and psoriasis: a population-based study. Ann Rheum Dis 2017;76(5):882–885.
4. Vallejo-Yagüe E, Burkard T, Möller B, et al. Comparison of psoriatic arthritis and rheumatoid arthritis patients across body mass index categories in Switzerland. J Clin Med 2021;10(14):3194.
5. Sun X, Du T. Trends in weight change patterns across life course among US adults, 1988-2018: population-based study. BMC Public Health 2023;23(1):2168.
6. Baker JF, Reed G, Kremer J. Weight fluctuation and the risk of cardiovascular events in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2022;74(2):229–235.
7. McMenamin KJ, Harris TB, Baker JF. Weight variability, physical functioning and incident disability in older adults. J Cachexia Sarcopenia Muscle 2023;14(4):1648–1656.
8. Bangalore S, Fayyad R, DeMicco DA, et al. Body weight variability and cardiovascular outcomes in patients with type 2 diabetes mellitus. Circ Cardiovasc Qual Outcomes 2018;11(11):e004724.
9. Cleary MP, Jacobson MK, Phillips FC, et al. Weight-cycling decreases incidence and increases latency of mammary tumors to a greater extent than does chronic caloric restriction in mouse mammary tumor virus-transforming growth factor-alpha female mice. Cancer Epidemiol Biomarkers Prev 2002;11(9):836–843.
10. Li X, Jiang L, Yang M, et al. Impact of weight cycling on CTRP3 expression, adipose tissue inflammation and insulin sensitivity in C57BL/6J mice. Exp Ther Med 2018;16(3):2052–2059.
11. Giles JT, Allison M, Blumenthal RS, et al. Abdominal adiposity in rheumatoid arthritis: association with cardiometabolic risk factors and disease characteristics. Arthritis Rheum 2010;62(11):3173–3182.
12. Letarouilly JG, Flipo RM, Cortet B, et al. Body composition in patients with rheumatoid arthritis: a narrative literature review. Ther Adv Musculoskelet Dis 2021;13:1759720X211015006.
13. Carvalho GD, Bonfiglioli K, Caparbo VF, et al. Changes to body composition in women with long-standing established rheumatoid arthritis: differences by level of disease activity. J Clin Densitom 2020;23(4):639–646.
14. Baker JF, Harris T, Rapoport A, et al. Validation of a description of sarcopenic obesity defined as excess adiposity and low lean mass relative to adiposity. J Cachexia Sarcopenia Muscle 2020;11(6):1580–1589.
15. Weber D, Long J, Leonard MB, et al. Development of novel methods to define deficits in appendicular lean mass relative to fat mass. PLoS One 2016;11(10):e0164385.
16. Baker JF, Giles JT, Weber D, et al. Sarcopenic obesity in rheumatoid arthritis: prevalence and impact on physical functioning. Rheumatology (Oxford) 2022;61(6):2285–2294.
17. Baker JF, Von Feldt J, Mostoufi-Moab S, et al. Deficits in muscle mass, muscle density, and modified associations with fat in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2014;66(11):1612–1618.
18. Baker JF, Mostoufi-Moab S, Long J, et al. Intramuscular fat accumulation and associations with body composition, strength, and physical functioning in patients with rheumatoid arthritis. Arthritis Care Res (Hoboken) 2018;70(12):1727–1734.
19. Khoja SS, Moore CG, Goodpaster BH, et al. Skeletal muscle fat and its association with physical function in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2018;70(3):333–342.
20. Baker JF, Billig E, Michaud K, et al. Weight loss, the obesity paradox, and the risk of death in rheumatoid arthritis. Arthritis Rheumatol 2015;67(7):1711–1717.
21. Rohm TV, Meier DT, Olefsky JM, et al. Inflammation in obesity, diabetes, and related disorders. Immunity 2022;55(1):31–55.
22. Nicolau J, Lequerré T, Bacquet H, et al. Rheumatoid arthritis, insulin resistance, and diabetes. Joint Bone Spine 2017;84(4):411–416.
23. Ristić GG, Subota V, Stanisavljević D, et al. Impact of disease activity on impaired glucose metabolism in patients with rheumatoid arthritis. Arthritis Res Ther 2021;23(1):95.
24. Manrique-Arija S, Ureña I, Valdivielso P, et al. Insulin resistance and levels of adipokines in patients with untreated early rheumatoid arthritis. Clin Rheumatol 2016;35(1):43–53.
25. Arias de la Rosa I, Escudero-Contreras A, Rodríguez-Cuenca S, et al. Defective glucose and lipid metabolism in rheumatoid arthritis is

determined by chronic inflammation in metabolic tissues. J Intern Med 2018;284(1):61–77.

26. Gallagher L, Cregan S, Biniecka M, et al. Insulin-resistant pathways are associated with disease activity in rheumatoid arthritis and are subject to disease modification through metabolic reprogramming: a potential novel therapeutic approach. Arthritis Rheumatol 2020; 72(6):896–902.
27. Li J, Chen Y, Liu Q, et al. Mechanistic and therapeutic links between rheumatoid arthritis and diabetes mellitus. Clin Exp Med 2023;23(2): 287–299.
28. Mohallem R, Aryal UK. Regulators of TNFα mediated insulin resistance elucidated by quantitative proteomics. Sci Rep 2020;10(1): 20878.
29. Tian Z, Mclaughlin J, Verma A, et al. The relationship between rheumatoid arthritis and diabetes mellitus: a systematic review and meta-analysis. Cardiovasc Endocrinol Metab 2021;10(2):125–131.
30. Dubreuil M, Rho YH, Man A, et al. Diabetes incidence in psoriatic arthritis, psoriasis and rheumatoid arthritis: a UK population-based cohort study. Rheumatology (Oxford) 2014;53(2):346–352.
31. Su CC, Chen IC, Young FN, et al. Risk of diabetes in patients with rheumatoid arthritis: a 12-year retrospective cohort study. J Rheumatol 2013;40(9):1513–1518.
32. Jiang P, Li H, Li X. Diabetes mellitus risk factors in rheumatoid arthritis: a systematic review and meta-analysis. Clin Exp Rheumatol 2015; 33(1):115–121.
33. Chung CP, Oeser A, Solus JF, et al. Prevalence of the metabolic syndrome is increased in rheumatoid arthritis and is associated with coronary atherosclerosis. Atherosclerosis 2008;196(2):756–763.
34. Jin Y, Chen SK, Liu J, et al. Risk of incident type 2 diabetes mellitus among patients with rheumatoid arthritis: a population-based cohort study. Arthritis Care Res (Hoboken) 2020;72(9):1248–1256.
35. Baker JF, England BR, George M, et al. Disease activity, cytokines, chemokines and the risk of incident diabetes in rheumatoid arthritis. Ann Rheum Dis 2021;80(5):566–572.
36. Di Muzio C, Cipriani P, Ruscitti P. Rheumatoid arthritis treatment options and type 2 diabetes: unravelling the association. BioDrugs 2022;36(6):673–685.
37. Solomon DH, Massarotti E, Garg R, et al. Association between disease-modifying antirheumatic drugs and diabetes risk in patients with rheumatoid arthritis and psoriasis. JAMA 2011;305(24):2525–2531.
38. Lim WS, Teoh SE, Tang ASP, et al. The effects of anti-TNF-α biologics on insulin resistance and insulin sensitivity in patients with rheumatoid arthritis: an update systematic review and meta-analysis. Diabetes Metab Syndr 2024;18(4):103001.
39. Ruscitti P, Masedu F, Alvaro S, et al. Anti-interleukin-1 treatment in patients with rheumatoid arthritis and type 2 diabetes (TRACK): a multicentre, open-label, randomised controlled trial. PLoS Med 2019; 16(9):e1002901.
40. Ruscitti P, Berardicurti O, Cipriani P, et al; TRACK study group. Benefits of anakinra versus TNF inhibitors in rheumatoid arthritis and type 2 diabetes: long-term findings from participants furtherly followed-up in the TRACK study, a multicentre, open-label, randomised, controlled trial. Clin Exp Rheumatol 2021;39(2):403–406.
41. Everett BM, Donath MY, Pradhan AD, et al. Anti-inflammatory therapy with canakinumab for the prevention and management of diabetes. J Am Coll Cardiol 2018;71(21):2392–2401.
42. Genovese MC, Burmester GR, Hagino O, et al. Interleukin-6 receptor blockade or TNFα inhibition for reducing glycaemia in patients with RA and diabetes: post hoc analyses of three randomised, controlled trials. Arthritis Res Ther 2020;22(1):206.
43. Baker JF, Reed G, Mikuls TR, et al. Metabolic syndrome adipokines, and response to advanced therapies in rheumatoid arthritis. Arthritis Rheumatol 2025;77(3):263–271.
44. Moroni L, Farina N, Dagna L. Obesity and its role in the management of rheumatoid and psoriatic arthritis. Clin Rheumatol 2020;39(4): 1039–1047.
45. van Halm VP, Peters MJL, Voskuyl AE, et al. Rheumatoid arthritis versus diabetes as a risk factor for cardiovascular disease: a cross-sectional study, the CARRE Investigation. Ann Rheum Dis 2009; 68(9):1395–1400.
46. Ridker PM, Everett BM, Thuren T, et al; CANTOS Trial Group. antiinflammatory therapy with canakinumab for atherosclerotic disease. N Engl J Med 2017;377(12):1119–1131.
47. Charles-Schoeman C, Gonzalez-Gay MA, Kaplan I, et al. Effects of tofacitinib and other DMARDs on lipid profiles in rheumatoid arthritis: implications for the rheumatologist. Semin Arthritis Rheum 2016; 46(1):71–80.
48. Myasoedova E, Crowson CS, Kremers HM, et al. Total cholesterol and LDL levels decrease before rheumatoid arthritis. Ann Rheum Dis 2010;69(7):1310–1314.
49. Karpouzas GA, Ormseth SR, Ronda N, et al. Lipoprotein oxidation may underlie the paradoxical association of low cholesterol with coronary atherosclerotic risk in rheumatoid arthritis. J Autoimmun 2022; 129:102815.
50. Charles-Schoeman C, Fleischmann R, Davignon J, et al. Potential mechanisms leading to the abnormal lipid profile in patients with rheumatoid arthritis versus healthy volunteers and reversal by tofacitinib. Arthritis Rheumatol 2015;67(3):616–625.
51. Song H, Saito K, Fujigaki S, et al. IL-1β and TNF-α suppress apolipoprotein (apo) E secretion and apo A-I expression in HepG2 cells. Cytokine 1998;10(4):275–280.
52. Watanabe J, Charles-Schoeman C, Miao Y, et al. Proteomic profiling following immunoaffinity capture of high-density lipoprotein: association of acute-phase proteins and complement factors with proinflammatory high-density lipoprotein in rheumatoid arthritis. Arthritis Rheum 2012;64(6):1828–1837.
53. Charles-Schoeman C, Hyde C, Guan S, et al. Relationship between paraoxonase-1 genotype and activity, and major adverse cardiovascular events and malignancies in patients with rheumatoid arthritis receiving tofacitinib. J Rheumatol 2023;50(12):1573–1580.
54. Charles-Schoeman C, Wang J, Shahbazian A, et al. Suppression of inflammatory arthritis in human serum paraoxonase 1 transgenic mice. Sci Rep 2020;10(1):16848.
55. Robertson J, Porter D, Sattar N, et al. Interleukin-6 blockade raises LDL via reduced catabolism rather than via increased synthesis: a cytokine-specific mechanism for cholesterol changes in rheumatoid arthritis. Ann Rheum Dis 2017;76(11):1949–1952.
56. Myasoedova E, Chandran A, Ilhan B, et al. The role of rheumatoid arthritis (RA) flare and cumulative burden of RA severity in the risk of cardiovascular disease. Ann Rheum Dis 2016;75(3):560–565.
57. Ajeganova S, Andersson MLE, Frostegård J, et al. Disease factors in early rheumatoid arthritis are associated with differential risks for cardiovascular events and mortality depending on age at onset: a 10-year observational cohort study. J Rheumatol 2013;40(12):1958–1966.
58. Ljung L, Rantapää-Dahlqvist S, Jacobsson LTH, et al. Response to biological treatment and subsequent risk of coronary events in rheumatoid arthritis. Ann Rheum Dis 2016;75(12):2087–2094.
59. Agca R, Heslinga SC, Rollefstad S, et al. EULAR recommendations for cardiovascular disease risk management in patients with rheumatoid arthritis and other forms of inflammatory joint disorders: 2015/2016 update. Ann Rheum Dis 2017;76(1):17–28.

60. Solomon DH, Giles JT, Liao KP, et al; TARGET Trial Consortium. Reducing cardiovascular risk with immunomodulators: a randomised active comparator trial among patients with rheumatoid arthritis. Ann Rheum Dis 2023;82(3):324–330.
61. Singh S, Fumery M, Singh AG, et al. Comparative risk of cardiovascular events with biologic and synthetic disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2020;72(4):561–576.
62. Souto A, Salgado E, Maneiro JR, et al. Lipid profile changes in patients with chronic inflammatory arthritis treated with biologic agents and tofacitinib in randomized clinical trials: a systematic review and meta-analysis. Arthritis Rheumatol 2015;67(1):117–127.
63. Giles JT, Sattar N, Gabriel S, et al. Cardiovascular safety of tocilizumab versus etanercept in rheumatoid arthritis: a randomized controlled trial. Arthritis Rheumatol 2020;72(1):31–40.
64. Ytterberg SR, Bhatt DL, Mikuls TR, et al; ORAL Surveillance Investigators. Cardiovascular and cancer risk with tofacitinib in rheumatoid arthritis. N Engl J Med 2022;386(4):316–326.
65. Szekanecz Z, Buch MH, Charles-Schoeman C, et al. Efficacy and safety of JAK inhibitors in rheumatoid arthritis: update for the practising clinician. Nat Rev Rheumatol 2024;20(2):101–115.
66. Kitas GD, Nightingale P, Armitage J, et al; TRACE RA Consortium. A multicenter, randomized, placebo-controlled trial of atorvastatin for the primary prevention of cardiovascular events in patients with rheumatoid arthritis. Arthritis Rheumatol 2019;71(9):1437–1449.
67. Crowson CS, Gabriel SE, Semb AG, et al; Trans-Atlantic Cardiovascular Consortium for Rheumatoid Arthritis. Rheumatoid arthritis-specific cardiovascular risk scores are not superior to general risk scores: a validation analysis of patients from seven countries. Rheumatology (Oxford) 2017;56(7):1102–1110.
68. Giles JT, Wasko MCM, Chung CP, et al. Exploring the lipid paradox theory in rheumatoid arthritis: associations of low circulating low-density lipoprotein concentration with subclinical coronary atherosclerosis. Arthritis Rheumatol 2019;71(9):1426–1436.
69. Chen CI, Wang L, Wei W, et al. Burden of rheumatoid arthritis among US Medicare population: co-morbidities, health-care resource utilization and costs. Rheumatol Adv Pract 2018;2(1):rky005.
70. Protogerou AD, Panagiotakos DB, Zampeli E, et al. Arterial hypertension assessed "out-of-office" in a contemporary cohort of rheumatoid arthritis patients free of cardiovascular disease is characterized by high prevalence, low awareness, poor control and increased vascular damage-associated "white coat" phenomenon. Arthritis Res Ther 2013;15(5):R142.
71. Sung KC, Suh JY, Kim BS, et al. High sensitivity C-reactive protein as an independent risk factor for essential hypertension. Am J Hypertens 2003;16(6):429–433.
72. Pérez-De-Lis M, Akasbi M, Sisó A, et al. Cardiovascular risk factors in primary Sjögren's syndrome: a case-control study in 624 patients. Lupus 2010;19(8):941–948.
73. Ballocca F, D'Ascenzo F, Moretti C, et al. Predictors of cardiovascular events in patients with systemic lupus erythematosus (SLE): a systematic review and meta-analysis. Eur J Prev Cardiol 2015;22(11):1435–1441.
74. Bartoloni E, Alunno A, Gerli R. Hypertension as a cardiovascular risk factor in autoimmune rheumatic diseases. Nat Rev Cardiol 2018;15(1):33–44.
75. Loperena R, Harrison DG. Oxidative stress and hypertensive diseases. Med Clin North Am 2017;101(1):169–193.
76. Rothman AM, MacFadyen J, Thuren T, et al. Effects of interleukin-1β inhibition on blood pressure, incident hypertension, and residual inflammatory risk: a secondary analysis of CANTOS. Hypertension 2020;75(2):477–482.
77. Ormseth MJ, Oeser AM, Chung CP, et al. Ambulatory Blood pressure in patients with rheumatoid arthritis: association with immune activation. J Rheumatol 2024;51(9):870–876.
78. Anyfanti P, Gavriilaki E, Douma S, et al. Endothelial dysfunction in patients with rheumatoid arthritis: the role of hypertension. Curr Hypertens Rep 2020;22(8):56.
79. Provan SA, Semb AG, Hisdal J, et al. Remission is the goal for cardiovascular risk management in patients with rheumatoid arthritis: a cross-sectional comparative study. Ann Rheum Dis 2011;70(5):812–817.
80. Bergholm R, Leirisalo-Repo M, Vehkavaara S, et al. Impaired responsiveness to NO in newly diagnosed patients with rheumatoid arthritis. Arterioscler Thromb Vasc Biol 2002;22(10):1637–1641.
81. Baker JF, Sauer B, Teng CC, et al. Initiation of disease-modifying therapies in rheumatoid arthritis is associated with changes in blood pressure. J Clin Rheumatol 2018;24(4):203–209.
82. Hadwen B, Stranges S, Barra L. Risk factors for hypertension in rheumatoid arthritis patients–a systematic review. Autoimmun Rev 2021;20(4):102786.
83. Mangoni AA, Tommasi S, Zinellu A, et al. Methotrexate and vasculoprotection: mechanistic insights and potential therapeutic applications in old age. Curr Pharm Des 2019;25(39):4175–4184.
84. Wolfe F, Michaud K. The risk of myocardial infarction and pharmacologic and nonpharmacologic myocardial infarction predictors in rheumatoid arthritis: a cohort and nested case-control analysis. Arthritis Rheum 2008;58(9):2612–2621.
85. Costello RE, Yimer BB, Roads P, et al. Glucocorticoid use is associated with an increased risk of hypertension. Rheumatology (Oxford) 2021;60(1):132–139.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1175–1179
DOI 10.1002/acr.25521

American College of Rheumatology
Empowering Rheumatology Professionals

EDITORIAL

Ageism in Rheumatologic Care: Ensuring Equity and Quality for Older Adults

Jiha Lee,[1] Devyani Misra,[2] and Una E. Makris[3]

The opening line of most medical documentation often reads, "Patient is a [xx]-year-old presenting for evaluation of…" This standard practice of categorizing patients primarily by their age shapes our perception, subtly perpetuating a divisive labeling approach. In rheumatology, patient age is as pivotal as identifying erosive changes when determining treatment strategies for rheumatoid arthritis (RA).[1] However, there is a concerning trend: the older the patient, the less likely they are to receive disease-modifying antirheumatic drugs (DMARDs), even when they are functionally and cognitively capable and would benefit from the standard of care.[1,2] This may represent ageism.

The World Health Organization (WHO) defines ageism as stereotyping (how we think), prejudice (how we feel), and discrimination (how we act) directed toward individuals based on their age. Ageism is widespread in our society and can manifest in several ways.[3] It can be internalized, leading to negative feelings about one's own aging process. It can be interpersonal, when people treat others differently based on age, sometimes through careless comments or deliberate actions. Additionally, ageism can be structural, when discrimination against older adults is embedded within institutions, policies, and practices. The psychosocial impacts of ageism include increased feelings of isolation, helplessness, and frustration. Older adults might feel their concerns are not taken seriously, leading to mistrust in the health care system and decreased adherence to treatment regimens.

Ageism magnifies negative health impacts, including reduced lifespan, prolonged disability, and accelerated cognitive decline.[4] Although ageism can affect people of all ages throughout their lives, the negative impacts on health are greater in older individuals.[3] Alarmingly, one in two people worldwide hold ageist attitudes toward older adults.[5] With the US population aging rapidly—those aged 65 years old and older are projected to rise from 17% in 2022 to 23% by 2050—the number of older adults living with rheumatologic diseases is also increasing. Older adults often face additional challenges such as limited mobility, chronic pain, and other geriatric syndromes (eg, falls and frailty). Inadequate treatment possibly rooted in ageism further exacerbates these poor health outcomes.

The economic burden of ageism on the US health care system is also substantial. A staggering $63 billion or 15% of total health care expenditure related to the eight most expensive health conditions affecting older adults was attributed to ageism, after adjusting for age and sex.[6] For musculoskeletal disorders alone, excessive costs of $2.1 billion were attributed to age discrimination and $4.4 billion to negative age stereotypes. This financial strain affects both patients and the health care system at large.

In this editorial, we aim to raise awareness among our colleagues about the pervasive yet often overlooked issue of ageism. The implications of ageism within the field of rheumatology need recognition. Although research from other fields offers valuable insights, we must focus more on understanding and addressing ageism in rheumatologic care from both a clinical and patient perspective. We hope this discussion, with the perspective from an older adult with rheumatic disease (Lawrence "Rick" Phillips), will illuminate the path toward better, more equitable care for all patients.

Supported by the National Institute on Aging, NIH (grant K23-AG-082727 awarded to Dr Lee). Dr Makris's work was supported in part by VA Health Services Research and Development (grant HX003350, IIR 20–256) and the NIH (grant NIA-P30-AG022845).

[1]Jiha Lee, MD, MHS: University of Michigan, Ann Arbor, Michigan; [2]Devyani Misra, MD, MS: Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts; [3]Una E. Makris, MD, MSc: University of Texas Southwestern Medical Center, Dallas, Texas, and VA North Texas Health Care System, Dallas, Texas.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25521.

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

Submitted for publication February 24, 2025; accepted February 25, 2025.

As a patient, Phillips noted that "ageism happens (in health care) when people treat older [adults] with limited opportunities based primarily on age." "A rheumatologist once asked how long I expected to live. He indicated that would help define what care was appropriate. [The rheumatologist] seemed perplexed when I said age 200. The thing is, he was wondering why I was asking about changing biologic medication. At my age (51), he was perplexed as to why I was asking about more aggressive treatment. Why? I asked because I do not want my joints to hurt. I asked if that was uncommon for people of any age. No, he said he supposed not." (Phillips)

Ageism in rheumatologic care: manifestation and detrimental effects

Ageism in rheumatologic care can manifest in various ways, often starting with stereotypes and misconceptions.[7] At the individual level, self-directed ageism can influence treatment choice by affecting older adults' perceptions and attitudes toward aggressiveness and/or long-term side effects of medications. Furthermore, internalized ageism can decrease participation in behaviors known to enhance health outcomes, such as exercise and mind–body practices.[8] Conversely, this may lead older adults to engage in risky health behaviors, including unhealthy eating, excessive drinking, or smoking, ultimately reducing their overall quality of life.

At the interpersonal level, health care provider bias critically translates into disparities in care. Although chronologic age alone does not provide a true indicator of an older adult's health status, it may deter the standard of care.[1] Among older adults with a new diagnosis of late-onset RA, less than a third were initiated on some form of DMARD despite their general safety and efficacy in this population.[2] Ageism may contribute to undertreatment because health care professionals with less favorable stereotypical beliefs regarding aging were more focused on the risks of medical intervention rather than their benefits when counseling older adults with rheumatic diseases.[9] Phillips emphasized, "I understand the options when they are offered. I cannot understand why some physicians want to make those decisions for me without my input. Doctors do not live in my experience. Give me the options, and I will decide how to proceed."

Another common but flawed belief is that pain and disability are natural parts of aging, resulting in the underdiagnosis and undertreatment of rheumatologic conditions in older adults.[10] This misconception can cause delayed diagnosis and treatment, leading to disease progression and a decreased quality of life. At the institutional or policy level, older adults are often excluded from clinical trials, leading to a dearth of evidence base for age-friendly care and clinical practice guidelines. A systematic review showed that a third of clinical trials involving patients with RA and osteoarthritis had arbitrary age cutoffs.[11] Addressing the underrepresentation of older adults in research was recognized as a priority, and the National Institutes of Health implemented a policy to require researchers to include individuals of all ages in clinical research.[12] The Targets, Team, Tools, Time, and Techniques (5Ts) framework enhances the inclusion of older adults in clinical research by addressing specific barriers and promoting a more inclusive approach.[13] By setting clear inclusion goals, assembling multidisciplinary teams, using age-sensitive tools, allocating sufficient time, and implementing tailored techniques, researchers can create a conducive environment that respects and accommodates the unique needs of older adults.

Time, communication, and complexity: factors exacerbating ageism in rheumatology

Health care professionals are often unable to spend sufficient time understanding the unique needs of older patients, which exacerbates existing issues related to ageism. Older patients may have hearing or vision impairments necessitating more time and effort to ensure effective communication. Cognitive impairment can also impact assessment and subsequent time it takes to communicate management approaches in a way that both the older patient and their caregiver(s) understand. Rushed consultations can lead to misunderstandings, challenged trust or therapeutic alliance, and poorer health outcomes.

Multimorbidity in older patients can lead to reduced effectiveness of DMARD therapy and atypical presentations of adverse drug effects.[14,15] Additionally, treatment goals and preferences may differ from those of younger adults, making a one-size-fits-all approach inappropriate and highlighting the need for a patient-centered and age-friendly approach.[16] Phillips believes "rheumatologists should see older patients as people first, patients second, and only then older mixed in with other things in our chart and history." Without thorough, individualized assessments, rheumatology health care professionals risk suboptimal treatment outcomes.

Emotional and psychological needs are often overlooked. Issues such as isolation, depression, or anxiety can profoundly impact older patients' well-being, requiring time, empathy, and interdisciplinary collaboration to address.[16] The connection and communication styles between health care professional(s) and their older patients are critical, especially, as Phillips pointed out, when navigating "the challenging business of interacting with doctors and other healthcare providers." The patient's care experience begins the moment they enter a practice, and as Phillips shared, "ageism starts at the front desk and ripples right through a practice." In today's world, in which electronic health records and portal messages are increasingly relied upon, the clinical staff play crucial roles as messengers and gatekeepers of information.

Phillips emphasized that effective communication from both office staff and clinicians is essential when caring for older adults: "Talking to me like I'm a child makes me want to avoid visiting the doctor altogether. I advise your staff to avoid losing patients out the back door based on their interaction at the front door." Phillips was referring to "elderspeak," a communication style that involves using simplified vocabulary, exaggerated intonation, and a condescending tone with older adults, often stemming from implicit ageism.[17] Although it may be used with good intentions, elderspeak can come across as patronizing and may undermine patient–provider trust, exacerbate social isolation, and lead to resistance to care.

Intersectionality of ageism with other "-isms"

Ageism is a significant social determinant of health, often intersecting with other forms of discrimination, such as racism and sexism, to exacerbate health care disparities, particularly for older adults with chronic conditions like rheumatic diseases.[18,19] Policies and practices reflecting bias systematically disadvantage marginalized groups while ageism perpetuates stereotypes and biases against older adults.[20] This intersection of discriminatory factors creates barriers to accessing quality health care. Theories like "double jeopardy" and "cumulative inequality" elucidate how the interplay of these forms of discrimination leads to compounded, detrimental effects on health outcomes.[21] For instance, older adults from minority backgrounds may encounter prejudices that lead to misdiagnoses, insufficient treatment, and mistrust in the health care system, resulting in more severe disease activity and progression.[22–24] Gender bias can also add another layer of discrimination because health care professionals might dismiss or underestimate the symptoms of rheumatic diseases in older women.[25,26] These overlapping -isms further restrict access to necessary medications, specialist consultations, and comprehensive management, ultimately leading to poorer health outcomes and increased chronic conditions. Addressing these issues effectively requires promoting comprehensive health care practices that ensure all older adults receive equitable, culturally competent care.[18,22,27]

Strategies to mitigate ageism in rheumatologic care

In the global report on ageism, the WHO has outlined strategies for combating ageism.[5] Education and awareness are key in combating ageism within health care and ourselves.[5]

At the patient level, providing older adults with resources and education about their rheumatic disease condition enables them to advocate effectively for themselves. Understanding their diagnosis and treatment options empowers them to engage in informed discussions with their health care team. Older adults with rheumatic diseases should be encouraged to share their functional goals, thereby opening up conversations with their rheumatology team about optimal treatments.

At the clinic level, rheumatology health care professional training programs should focus on recognizing and countering ageist attitudes. This involves avoiding broad generalizations about older adults and understanding that chronological age does not necessarily reflect physiologic age. Leading national and international organizations combating ageism emphasize the need for language and imagery free of assumptions and judgment. Terms that stigmatize or dramatize aging, such as "elderly," "senior," or "demographic cliff," should be avoided, as well as any portrayal of older adults as having less worth, such as suggestions of overusing health care.[28] Instead, as a health care community, we should provide a balanced perspective on both the opportunities and challenges presented by demographic changes. The Geriatric 5Ms (Medication, Mobility, Mind, Multicomplexity, What Matters Most) framework provides a holistic approach to understanding older adults as individuals with distinct complexities, functional abilities, psychosocial needs, and care objectives.[16,29] Rheumatology health care professionals who are knowledgeable about the 5Ms tend to have more favorable and nuanced views about aging.[9] Applying the Geriatric 5Ms to assess social determinants of health can also help identify how structural racism intersects with ageism.[29] By acquiring the knowledge and skills to tackle age-related discrimination and inequality, rheumatologists can provide more empathetic and personalized care.

Collaborative care from an interdisciplinary team including rheumatologists, geriatricians, pain management specialists, clinical psychologists, clinical pharmacists doctors, physical therapists, and occupational therapists can help create a comprehensive treatment plan. Such a holistic approach ensures that all aspects of an older patient's health are considered, including their preferences and goals of care, leading to more effective and personalized care.

> *"[With age,] increasingly, the doctors have to interact and do so seamlessly. I rely on that, and for the most part, that interaction requires me to remind each team member to react to other team members' notes. I am finding that doctors who cannot interact well with the rest of my team are more of a liability than an asset."* (Phillips)

At the institutional level, policy and systemic changes are necessary to institutionalize the fight against ageism. Advocating for health care policies that promote the inclusion of older adults

in research is essential to develop evidence-based, age-friendly/appropriate guidelines that can help ensure that age is not a barrier to receiving appropriate care. Moreover, older adults should be recognized as important stakeholders in the process of developing age-friendly health care systems to reduce "othering"—an imaginary boundary between our present and future selves—and improve interpersonal and intergenerational collaboration.

Phillips himself is an active patient advocate and has "written [Centers for Medicare and Medicaid Services] and asked that rheumatologists receive better reimbursement. [Because] a rheumatologist will not find any better supporters than their older client base."(Phillips)

Conclusions

Ageism in rheumatologic care poses significant challenges. By recognizing the unique needs of older adults, fostering interdisciplinary collaborations, implementing policy changes, and leveraging innovations in care, we can ensure that all patients, regardless of age, receive the equitable and quality care they deserve.

"I want to be in partnership with my care team. But they must understand they are my care team, and I will be the leader. Each of us has a role on that team. My role is to make decisions about my care, given the options. Their part is to give me the options and be sure I understand the risks. I employ them to help me understand the options and evaluate the risks."
(Phillips)

ACKNOWLEDGMENTS

The patient voice and quotes included in this featured article are with permission from Lawrence "Rick" Phillips. Phillips is 67 years old, living in Indiana, and is diagnosed with RA (since 2000), type 1 diabetes, and chronic kidney disease. He served as the patient perspective representative at our American College of Rheumatology Convergence Community Hub on Aging in 2021, speaking about ageism. We thank Phillips for his time, expertise, and willingness to work with us on this article.

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 Makris 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. Fraenkel L, Rabidou N, Dhar R. Are rheumatologists' treatment decisions influenced by patients' age? Rheumatology (Oxford) 2006;45(12):1555–1557.
2. Lee J, Martindale J, Makris UE, et al. Initiation of disease-modifying antirheumatic drugs in older medicare beneficiaries with new diagnosis of late-onset rheumatoid arthritis. ACR Open Rheumatol 2023;5(12):694–700.
3. Allen JO, Solway E, Kirch M, et al. Experiences of everyday ageism and the health of older US adults. JAMA Netw Open 2022;5(6):e2217240.
4. Chang ES, Kannoth S, Levy S, et al. Global reach of ageism on older persons' health: a systematic review. PloS One 2020;15(1):e0220857.
5. World Health Organization. Global Report on Ageism. World Health Organization; 2021. Accessed October 27, 2024. https://iris.who.int/bitstream/handle/10665/340208/9789240016866-eng.pdf?sequence=1.
6. Levy BR, Slade MD, Chang ES, et al. Ageism amplifies cost and prevalence of health conditions. Gerontologist 2018;60(1):174–181.
7. Misra D, Buehring B, Yung R, et al. Ageism and rheumatic diseases. Lancet Rheumatol 2024;6(12):e817–e819.
8. Thøgersen-Ntoumani C, Kritz M, Grunseit A, et al. Barriers and enablers of vigorous intermittent lifestyle physical activity (VILPA) in physically inactive adults: a focus group study. Int J Behav Nutr Phys Act 2023;20(1):78.
9. Smith A, Achanta P, Makris U. Ageism in rheumatology: the health care professional's perspective. Arthritis Rheumatol 2023;75-(suppl 9).
10. Makris UE, Abrams RC, Gurland B, et al. Management of persistent pain in the older patient: a clinical review. JAMA 2014;312(8):825–836.
11. 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(6):983–989.
12. Inclusion across the lifespan in human subjects research. National Institutes of Health. Updated February 27, 2025. Accessed October 27, 2024. https://grants.nih.gov/policy-and-compliance/policy-topics/inclusion/lifespan
13. 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(2):342–346.
14. England BR, Yun H, Chen L, et al. Influence of multimorbidity on new treatment initiation and achieving target disease activity thresholds in active rheumatoid arthritis: a cohort study using the rheumatology informatics system for effectiveness registry. Arthritis Care Res (Hoboken) 2023;75(2):231–239.
15. Lee J, Singh N, Gray SL, et al. Optimizing medication use in older adults with rheumatic musculoskeletal diseases: deprescribing as an approach when less may be more. ACR Open Rheumatol 2022;4(12):1031–1041.

16. Buehring B, van Onna M, Myasoedova E, et al. Understanding the multiple dimensions of ageing: 5Ms for the rheumatologist. Lancet Rheumatol 2024;6(12):e892–e902.
17. Shaw CA, Gordon JK. Understanding elderspeak: an evolutionary concept analysis. Innov Aging 2021;5(3):igab023.
18. Farrell TW, Hung WW, Unroe KT, et al. Exploring the intersection of structural racism and ageism in healthcare. J Am Geriatr Soc 2022; 70(12):3366–3377.
19. Chrisler JC, Barney A, Palatino B. Ageism can be hazardous to women's health: ageism, sexism, and stereotypes of older women in the healthcare system. J Soc Issues 2016;72(1):86–104.
20. Bastos JL. Structural racism is a driver of fragmented care and acute care use among people with rheumatic conditions. Lancet Rheumatol 2025;7:e5–e6.
21. Brown TH, Homan P. Structural racism and health stratification: connecting theory to measurement. J Health Soc Behav 2024;65(1): 141–160.
22. Yang S, Santacroce L, Collins JE, et al. Association of historical redlining and present-day racialised economic segregation with health-care utilisation among individuals with rheumatic conditions in Massachusetts and surrounding areas of the USA: a retrospective cohort study. Lancet Rheumatol 2025;7(1):e33–e43.
23. Olson RM, Feldman CH. A critical look at race-based practices in rheumatology guidelines. Arthritis Care Res (Hoboken) 2022;74(10): 1602–1607.
24. Feldman CH, Ramsey-Goldman R. Widening disparities among patients with rheumatic diseases in the COVID-19 era: an urgent call to action. Arthritis Rheumatol 2020;72(9):1409–1411.
25. Coffey CM, Davis JM III, Crowson CS. The impact of gender on time to rheumatoid arthritis classification: a retrospective analysis of a population-based cohort. Rheumatol Int 2019;39(12):2025–2030.
26. Marzo-Ortega H, Navarro-Compán V, Akar S, et al. The impact of gender and sex on diagnosis, treatment outcomes and health-related quality of life in patients with axial spondyloarthritis. Clin Rheumatol 2022;41(11):3573–3581.
27. Dubreuil M, Ferucci ED, El-Gabalawy H, et al. Enhancing equity in clinical research: a multifaceted proposal for spondyloarthritis. J Rheumatol 2024;51(12):1247–1253.
28. Lundebjerg NE, Trucil DE, Hammond EC, et al. When it comes to older adults, language matters: Journal of the American Geriatrics Society adopts modified American Medical Association style. J Am Geriatr Soc 2017;65(7):1386–1388.
29. Miller RK, Young M, Chippendale R, et al. Using the Geriatric 5Ms to teach structural and social determinants of health. J Am Geriatr Soc 2023;71(12):3967–3972.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1180–1183
DOI 10.1002/acr.25598

EDITORIAL

Still's Lung Disease: Time to Recognize This Complication in Adults

Lauren A. Henderson

Pediatric rheumatologists often encounter astonished families when young children are diagnosed with chronic inflammatory arthritis. Yes, children get arthritis too! In this issue of *Arthritis Care & Research*, Slabbynck et al[1] demonstrate that a similar refrain can apply to adults: lung involvement in Still's disease is not just for kids. In fact, severe pulmonary complications occur in adult-onset Still's disease (AOSD) and are likely underrecognized.

Increased recognition of lung disease in children with systemic juvenile idiopathic arthritis

As far back as the 1980s, scattered reports in the literature documented restrictive lung disease (LD) and pulmonary hypertension in patients with systemic juvenile idiopathic arthritis (sJIA) and AOSD.[2,3] It was not until the early 2000s that pediatric rheumatologists noticed a marked increase in cases. In 2013, Kimura et al identified 25 such children with sJIA and co-occurring pulmonary hypertension, interstitial LD (ILD), and/or pathology demonstrating pulmonary alveolar proteinosis (PAP) and endogenous lipoid pneumonia (ELP) spectrum of findings, an association eventually termed sJIA-LD.[4] In this report, 68% of patients with sJIA-LD died, sparking widespread alarm in the field. Mortality rates from sJIA-LD have been lower in subsequent reports, ranging from 7% to 58%, although this is still well above the expected fatality rate of sJIA without LD.[5–7] These findings highlight the gravity of an sJIA-LD diagnosis and the urgent need to better understand this emerging entity.

Shared clinical features of LD in children and adults with Still's disease

Slabbynck et al[1] describe a patient with AOSD who developed diffuse LD, with pathology confirming ELP. This adult displayed many of the same risk factors identified in children with sJIA who develop LD, including refractory disease requiring multiple biologic disease-modifying antirheumatic drugs (bDMARDs), recurrent macrophage activation syndrome (MAS), and high levels of plasma interleukin-18 (IL-18).[4–6,8,9] In patients with sJIA, younger age at disease onset, adverse reactions to bDMARDs, congenital heart disease, and trisomy 21 have also been linked with the development of LD.[5,8,10]

sJIA-LD is characterized by several unusual features that were also noted in the patient described by Slabbynck et al. The onset of LD is insidious, and patients typically do not spontaneously report respiratory symptoms during the early phases of lung involvement.[5,8] Astute clinicians may uncover subtle signs with a detailed review of systems, such as cough or mild dyspnea with vigorous exercise or an upper respiratory tract infection. In children, clubbing is sometimes the only finding to herald the LD diagnosis.[5] As in this case, symptomatic patients often have advanced disease. In retrospect, this individual had evidence of LD on chest imaging at her initial presentation. Once treatment for AOSD was initiated, she was largely asymptomatic from a respiratory standpoint for two years until she developed hypoxia. The radiographic findings in the reported patient and children with sJIA-LD are quite distinct from other forms of ILD and include pleural, septal, and peribronchovascular thickening; tree-in-bud and ground glass opacities; peripheral consolidation; crazy paving; and lymphadenopathy.[5,8] Perhaps the most defining feature of sJIA-LD is pathology consistent with PAP/ELP. As in this patient, the PAP and ELP spectrum of findings can include type II alveolar cell hyperplasia, foamy macrophages, and cholesterol clefts.[5,8] Some patients also have evidence of vasculopathy and a lymphocytic infiltrate.[8] In studies evaluating patients with sJIA, hereditary and autoimmune (granulocyte–macrophage colony-stimulating factor [GM-CSF], autoantibodies) causes of PAP have not been found.[5,8]

The presentation of LD in this patient with AOSD closely mirrors the pediatric experience with sJIA-LD, including shared risk factors, similar clinical presentations, and overlapping

Lauren A. Henderson, MD, MMSc: Boston Children's Hospital, Boston, Massachusetts.
Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25598.

Address correspondence via email to Lauren A. Henderson, MD, MMSc, at lauren.henderson@childrens.harvard.edu.
Submitted for publication June 9, 2025; accepted June 13, 2025.

radiographic findings. Most striking is the common feature of PAP/ELP, which is a highly unusual pathologic finding. Although more adults with AOSD and pulmonary involvement need to be studied, the terms "sJIA-LD" and "AOSD-LD" almost certainly describe the same entity. The continued use of different names to label the same phenomena artificially segregates this condition based on age at presentation; indeed, a recent international guideline panel recommended that the term "Still's disease" be applied to both sJIA and AOSD.[11] Thus, a unifying term, such as "Still's-LD," seems more appropriate.

The importance of pulmonary screening in at-risk patients of all ages with Still's disease

Because many patients with Still's-LD are initially asymptomatic, most will not come to medical attention until significant pulmonary damage has accumulated. At this point, the disease can be rapidly progressive and extremely difficult to treat. However, a significant proportion of patients do respond to an escalation of immunosuppression. As reported in this case, attaining full control of the Still's disease is essential. In a single-center prospective cohort of 41 patients with sJIA-LD, Huang et al found that many patients continued therapy with bDMARDs along with the addition of other immunomodulators, most commonly mycophenolate mofetil, emapalumab (interferon-γ monoclonal antibody), or a JAK inhibitor.[7] Many patients were also started on lung-directed therapies, such as antimicrobial prophylaxis, inhaled glucocorticoids, and inhaled beta agonists. In this cohort, there was improvement in sJIA clinical activity (66% of patients), sJIA biomarkers (71%), and clinical symptoms of LD (37%); however, lung imaging was often unchanged (40%). The survival rate for this cohort at a median follow-up of 2.9 years was 93%, although 37% of patients remained oxygen dependent. For severely affected patients with progressive disease, allogeneic hematopoietic stem cell transplantation (HSCT) is an option. In a recently published multicenter, retrospective study of 13 patients with sJIA-LD who underwent HSCT, nine were alive at last follow-up, all of whom had a complete clinical response and did not require immunomodulatory therapy or supplemental oxygen.[12] Thus, LD is a serious manifestation of Still's disease; however, it can be treated once it is recognized.

For the aforementioned reasons, there are efforts underway to establish a systematic approach to pulmonary screening in children with sJIA. To date, two single-center algorithms have been published, both of which identify potential risk factors that should trigger a pulmonary evaluation.[13,14] These "red flag" features broadly include evidence of high disease activity (persistent systemic symptoms, recurrent MAS, and elevated markers of inflammation), concerning respiratory symptoms or examination findings, drug hypersensitivity–like characteristics (pruritic rash, eosinophilia, and adverse drug reactions), and HLA–DRB1*15 positivity. Once identified, patients with sJIA considered at risk for LD are referred to pulmonary for a baseline evaluation, which may include a chest radiograph, pulmonary function tests with diffusion capacity testing, a six-minute walk test, and overnight pulse oximetry. Based on this initial testing, further diagnostic studies (eg, high-resolution computed tomography of the chest, bronchoscopy, and echocardiogram) may be performed. Although this approach is a start, multicenter and/or society-supported guidelines are needed, coupled with studies that test the efficacy of this risk stratification–based approach. For now, adult rheumatologists can leverage the existing pediatric literature and should strongly consider a pulmonary evaluation in patients with AOSD with risk factors for LD.

What is driving the increased incidence of LD in Still's disease?

This question is intensely debated in the field of pediatric rheumatology. It is clear that cases of LD are on the rise, which has coincided with more widespread use of bDMARDs. Inflammatory cytokines (IL-1β, IL-6, and IL-18) are central to the pathogenesis of sJIA, and medications that inhibit IL-1β and IL-6 have transformed this previously devastating disease to a condition with excellent outcomes. Indeed, first-line treatment with IL-1β and IL-6 inhibitors is now accepted as standard of care for patients with sJIA.[15] Given the temporal association between bDMARDs and LD in sJIA, there has been concern that this phenomena may be due to drug reaction with eosinophilia and systemic symptoms (DRESS). In support of this hypothesis, many patients with sJIA-LD have a history of eosinophilia and meet Registry for Severe Cutaneous Adverse Reactions (RegiSCAR) criteria for DRESS. There are HLA class I associations linked with some drugs that cause DRESS. In the case of sJIA, a class II HLA allele (HLA–DRB1*15:01) has been found at increased frequencies in sJIA patients with DRESS-like features and LD compared to those without such manifestations.[5,16] The implication is that the medications that have dramatically improved outcomes for patients with sJIA cannot be used in any child with drug hypersensitivity–like features or LD.

The DRESS hypothesis has been challenged for several reasons. First, applying the RegiSCAR criteria to diagnose DRESS in patients with an inflammatory disease characterized by fever and rash is highly problematic because many patients will meet the definition for probable DRESS simply through usual disease features. Eosinophilia is common in sJIA patients with and without LD, suggesting that type 2 immune responses may be intrinsic to the disease.[6,9] There is also some evidence to indicate that eosinophilia is a marker of severity in patients with sJIA. In one single-center cohort, 20% of patients with sJIA had eosinophilia before exposure to bDMARDs. Further, MAS that developed at disease onset and before treatment initiation was associated with a three-fold increase rate of eosinophilia on univariate analysis.[6]

Similarly, HLA–DRB1*15 alleles occur at higher frequencies in some sJIA cohorts than expected in the general population. These HLA alleles are found in patients exposed to bDMARDs who never develop eosinophilia or LD.[6,9,17] Some preliminary findings also suggest that this HLA type may be associated with refractory disease. Importantly, multiple cohorts have documented LD in patients with sJIA never exposed to bDMARDs.[4–6,18] The converse is also true: LD has not been reported in patients with other diseases treated with IL-1β and IL-6 inhibitors. The experience of clinicians who treat these patients regularly is that stopping bDMARDs in active sJIA is disastrous. As reported in this case and the literature, many patients improve while continuing to take bDMARDs at escalated doses or in conjunction with other agents.[7] Accordingly, a recent consensus group recommended against discontinuing bDMARD therapy as part of treatment for LD.[11]

To reconcile these observations, Binstadt and Nigrovic postulated that the increased use of first-line IL-1β and IL-6 inhibitors in sJIA without cotreatment with glucocorticoids and/or conventional disease-modifying antirheumatic drugs may drive cytokine skewing that favors type 2 immune responses.[19] Unlike the DRESS theory, this cytokine plasticity hypothesis implies that bDMARDs can be continued in patients with sJIA with hypersensitivity-like features and LD, although potentially with the addition of other agents that target the immune system more broadly, such as glucocorticoids, methotrexate, calcineurin inhibitors, or JAK inhibitors. Although appealing, the cytokine plasticity hypothesis remains untested, and further studies are needed to support its validity.

Conclusions

As demonstrated by Slabbynck et al, lung involvement is a serious complication of Still's disease that can occur in adults. Early recognition coupled with treatment intensification to gain full disease control is essential to prevent irreversible pulmonary damage and death. The biologic basis of Still's-LD remains enigmatic, and collaborative efforts between pediatric and adult rheumatologists are needed to tackle this perplexing problem.

AUTHOR CONTRIBUTIONS

The author 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 Henderson confirms that the author has 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. Slabbynck K, Van Dorpe J, Peene I, et al. Endogenous lipoid pneumonia in adult autoinflammatory disease. Arthritis Care Res (Hoboken) 2025;77(10):1184–1193.
2. Corbett AJ, Zizic TM, Stevens MB. Adult-onset Still's disease with an associated severe restrictive pulmonary defect: a case report. Ann Rheum Dis 1983;42(4):452–454.
3. Padeh S, Laxer RM, Silver MM, et al. Primary pulmonary hypertension in a patient with systemic-onset juvenile arthritis. Arthritis Rheum 1991;34(12):1575–1579.
4. Kimura Y, Weiss JE, Haroldson KL, et al; Childhood Arthritis Rheumatology Research Alliance Carra Net Investigators. Pulmonary hypertension and other potentially fatal pulmonary complications in systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2013;65(5):745–752.
5. Saper VE, Chen G, Deutsch GH, et al; Childhood Arthritis and Rheumatology Research Alliance Registry Investigators. Emergent high fatality lung disease in systemic juvenile arthritis. Ann Rheum Dis 2019;78(12):1722–1731.
6. Wobma H, Arvila SR, Taylor ML, et al. Incidence and risk factors for eosinophilia and lung disease in biologic-exposed children with systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2023;75(10):2063–2072.
7. Huang Y, Sompii-Montgomery L, Patti J, et al. Disease course, treatments, and outcomes of children with systemic juvenile idiopathic arthritis-associated lung disease. Arthritis Care Res (Hoboken) 2024;76(3):328–339.
8. Schulert GS, Yasin S, Carey B, et al. Systemic juvenile idiopathic arthritis-associated lung disease: characterization and risk factors. Arthritis Rheumatol 2019;71(11):1943–1954.
9. Lerman AM, Mahmud SA, Alfath Z, et al. HLA-DRB1*15 and eosinophilia are common among patients with systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2023;75(10):2082–2087.
10. Aires BP, Wobma H, Samad A, et al. Severe features of systemic juvenile idiopathic arthritis in patients with congenital heart disease. J Rheumatol 2024;51(8):811–817.
11. Fautrel B, Mitrovic S, De Matteis A, et al. EULAR/PReS recommendations for the diagnosis and management of Still's disease, comprising systemic juvenile idiopathic arthritis and adult-onset Still's disease. Ann Rheum Dis 2024;83(12):1614–1627.
12. Matt MG, Drozdov D, Bendstrup E, et al. Allogeneic haematopoietic stem-cell transplantation for children with refractory systemic juvenile idiopathic arthritis and associated lung disease: outcomes from an international, retrospective cohort study. Lancet Rheumatol 2025;7(4):e243–e251.https://doi.org/10.1016/S2665-9913(24)00275-3
13. Wobma H, Bachrach R, Farrell J, et al. Development of a screening algorithm for lung disease in systemic juvenile idiopathic arthritis. ACR Open Rheumatol 2023;5(10):556–562.
14. Towe C, Grom AA, Schulert GS. Diagnosis and management of the systemic juvenile idiopathic arthritis patient with emerging lung disease. Paediatr Drugs 2023;25(6):649–658.
15. 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(4):521–537.
16. Saper VE, Ombrello MJ, Tremoulet AH, et al; Drug Hypersensitivity Consortium; INCHARGE Consortium. Severe delayed hypersensitivity reactions to IL-1 and IL-6 inhibitors link to common HLA-DRB1*15 alleles. Ann Rheum Dis 2022;81(3):406–415.

17. Erkens RGA, Calis JJA, Verwoerd A, et al. Recombinant interleukin-1 receptor antagonist is an effective first-line treatment strategy in new-onset systemic juvenile idiopathic arthritis, irrespective of HLA-DRB1 background and IL1RN variants. Arthritis Rheumatol 2024;76(1): 119–129.
18. Zhan W, Yang J, Qiu L, et al. Clinical characteristics and prognosis of interstitial lung disease in systemic juvenile idiopathic arthritis: a two-center retrospective observational cohort study. Pediatr Rheumatol Online J 2024;22(1):96.https://doi.org/10.1186/s12969-024-01028-5
19. Binstadt BA, Nigrovic PA. The conundrum of lung disease and drug hypersensitivity-like reactions in systemic juvenile idiopathic arthritis. Arthritis Rheumatol 2022;74(7):1122–1131.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1184–1193
DOI 10.1002/acr.25570

CLINICOPATHOLOGIC CONFERENCE

Endogenous Lipoid Pneumonia in Adult Autoinflammatory Disease

Katrien Slabbynck,[1] Jo Van Dorpe,[2] Isabelle Peene,[3] Karin Melsens,[4] Cedric Bosteels,[2] Rik Joos,[5] Dirk Elewaut,[6] Peggy Jacques,[6] Hans Slabbynck,[5] and Gaëlle Varkas[6]

We describe one of the first cases of endogenous lipoid pneumonia (ELP) in an adult patient with the clinical picture of adult-onset Still disease (AOSD) and a low penetrance genetic background of tumor necrosis factor receptor-associated periodic syndrome (TRAPS). This case highlights the complex pathophysiology of lung involvement in autoinflammatory diseases, operating at the interface of the innate and adaptive immune system.

This case presents a 53-year-old immunocompromised woman with treatment refractory autoinflammatory disease and history of macrophage activation syndrome (MAS), presenting to the emergency room with progressive dyspnea and fever. Upon evaluation, chest computed tomography showed diffuse lung disease. Extensive workup, including bronchoscopy with bronchoalveolar lavage, remained negative. Lung biopsy revealed an ELP with intra-alveolar accumulation of cholesterol crystals and foamy macrophages. In the years preceding the event, her autoinflammatory disease had shown to be refractory to both conventional systemic disease-modifying antirheumatic drugs and biologic treatments including tocilizumab, anakinra, and canakinumab. Because of new onset respiratory failure in the context of uncontrolled inflammation, after exclusion of infectious origin, pulse doses of systemic glucocorticoids were administered before induction with cyclophosphamide, followed by maintenance therapy with tacrolimus. Upon treatment, our patient recovered but retained severe interstitial lung disease. Only one case of ELP in adult autoinflammatory disease has been depicted in a patient diagnosed with AOSD, although the entity is more recognized in pediatric literature on systemic onset juvenile idiopathic arthritis (soJIA).

CASE PRESENTATION

History of present illness

A 53-year-old immunocompromised female patient, previously diagnosed with treatment refractory adult-onset Still disease (AOSD) and recurrent macrophage activation syndrome (MAS) two years earlier, presented to the emergency room with progressive dyspnea and fever.

Past medical history

Two years before the present admission, the patient was diagnosed with an autoinflammatory syndrome after being hospitalized for suspected *Chlamydia psittacosis* pneumonia. Her initial presentation consisted of progressive dyspnea, fever, sore throat, myalgia, and muscle weakness. The aforementioned diagnosis was considered due to close contact with canaries and a single mildly positive polymerase chain

[1]Katrien Slabbynck, MD: Ghent University, Ghent, Belgium; [2]Jo Van Dorpe, MD, PhD, Cedric Bosteels, MD, PhD: Ghent University and Ghent University Hospital, Ghent, Belgium; [3]Isabelle Peene, MD, PhD: Ghent University and Ghent University Hospital, Ghent, Center for Inflammation Research, Vlaams Insituut voor Biotechnologie (VIB)-Universiteit Gent, Zwijnaarde, and Algemeen Ziekenhuis (AZ) Sint-Jan, Brugge, Belgium; [4]Karin Melsens, MD, PhD: Universitair Ziekenhuis Brussel - Vrije Universiteit Brussel, Brussels, Belgium; [5]Rik Joos, MD, Hans Slabbynck, MD: Ziekenhuis Aan de Stroom (ZAS), Antwerpen, Belgium; [6]Dirk Elewaut, MD, PhD, Peggy Jacques, MD, PhD, Gaëlle Varkas, MD, PhD: Ghent University and Ghent University Hospital, Ghent, and Center for Inflammation Research, Vlaams Instituut voor Biotechnologie (VIB)-Universiteit Gent, Zwijnaarde, Belgium.

Data availability: Data can be made available upon reasonable request.

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.25570).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25570.

Address correspondence via email to Gaëlle Varkas, MD, PhD, at gaelle.varkas@ugent.be.

Submitted for publication September 23, 2024; accepted in revised form May 1, 2025.

reaction (PCR) swab, after other pathogens had been excluded. After several weeks in the intensive care unit (ICU), the diagnosis of autoinflammation with the clinical picture of AOSD was made by the rheumatologist based on prolonged spiking fever, sore throat, oligoarticular arthritis, mildly itchy maculopapular rash, hepatosplenomegaly, and diffuse adenopathy. At the time of diagnosis, the patient exhibited no residual respiratory symptoms nor clubbing. Laboratory results showed elevated parameters of inflammation, with ferritin up to 134,455 μg/L and raised liver enzymes, whereas blood eosinophil results were normal. A chest computed tomography (CT) performed during the workup revealed bilateral pleural effusions in the lower lobes and thickened interlobular apical septa, accompanied by minimal ground-glass opacities. There were no definitive signs of pulmonary infection. Later on, genetic testing for periodic fever syndromes revealed a low penetrance variant of unknown significance in the TNFRSF1A gene (c.362G>A(p.Arg121Gln) (R92Q)), a mutation observed in tumor necrosis factor receptor-associated periodic syndrome (TRAPS). Nevertheless, this variant is relatively common in European populations and considered a nonconfirmatory genotype for TRAPS.[1] Adult-onset TRAPS accounts for approximately 22% to 45% of cases and less commonly presents with abdominal pain or orbital edema in addition to systemic symptoms.[2, 3]

Apart from endometriosis, the patient's personal medical history up until her current diagnosis was unremarkable. There was no family history of autoimmune disease or of rheumatic musculoskeletal disorders or pulmonary conditions. She had never smoked, used drugs or medication, or had any occupational exposure to hazardous agents. She denied any recent travel.

Subsequent to her diagnosis, the patient responded well to high-dose systemic glucocorticoids. However, tapering proved challenging despite sequential association of nonsteroidal anti-inflammatory drugs, methotrexate, colchicine, and several biologic treatments including tocilizumab, anakinra, and canakinumab. Although there was a therapeutic response to anakinra with resolution of inflammatory symptoms for several months, she displayed secondary inefficacy despite dose escalation up to 200 mg per day. During her disease course, several episodes of suspected MAS with spiking fever, rash, neurologic symptoms, diarrhea, hypotension, anemia, elevated parameters of inflammation, hyperferritinemia, and high soluble CD25 developed. Serum cytokine analysis showed strong elevation of interleukin (IL)-1RA, IL-18 (up to 2,616.2 pg/mL [normal range <50 pg/mL]), IL-6, and CXCL9 (up to 2,908.1 pg/mL [normal range 150–550 pg/mL]). However, results from bone marrow and lymph node biopsies were negative for hemophagocytosis, whereas the skin biopsy was consistent with neutrophilic urticarial dermatitis showing hemophagocytosis.

Review of systems

The patient presented to the emergency room with exhaustion, headache, progressive dyspnea, diarrhea, spiking fever, and inflammatory myalgia and arthralgia. Further review of systems revealed no abnormalities, more specifically no skin rash or arthritis. The patient was currently treated with prednisolone 15 mg daily and canakinumab 300 mg every four weeks for refractory autoinflammatory disease.

Physical examination

Upon admission, the patient presented with a temperature of 40°C, a pulse rate of 147 beats per minute, blood pressure of 95/60 mm Hg, and oxygen saturation of 96% on 5 L of oxygen. The patient was tachypneic, with lung auscultation revealing bilateral fine crackles. On cardiac examination, no murmurs were detected. There was no evidence of arthritis, skin rash, or clubbing. Blood gas analysis confirmed hypoxemia with hypocapnia.

Laboratory evaluation

Laboratory examination included a hemoglobin level of 10.9 g/dL (normal range 11.7–15.1), leukocytes of 15.9 × 10^3/μL (normal range 4.30–9.64 × 10^3/μL) with 12,860/μL neutrophils, blood eosinophils within normal range, platelets of 281 × 10^3/μL (normal range 175–343 × 10^3/μL). C-reactive protein (CRP) was elevated up to 213.7 mg/L (normal range <5 mg/L), with procalcitonin and ferritin, respectively, up to 25.2 ng/mL (normal range <0.1 ng/mL) and 14,579 μg/L (normal range <250 μg/L). Liver function was normal, as was the lipid profile including triglycerides. Creatinine was elevated, which was attributed to dehydration due to fever, reduced intake, and diarrhea.

Conventional chest-CT indicated bilateral peribronchovascular consolidations and multiple paratracheal, subcarinal, and hilar adenopathies, with signs of congestion and limited unilateral pleural fluid. Furthermore, high-resolution chest-CT (HRCT) showed diffuse lung disease (LD), predominant in the lower and middle lobes with septal thickening involving the periphery of several lobes, some adjacent ground-glass opacities, bilateral peribronchovascular thickening, and some pleural thickening (Figure 1). Upon review of previous imaging obtained at the time of the AOSD diagnosis two years prior, chest-CT already demonstrated limited septal thickening in the apical lobes and some ground-glass opacities, albeit less pronounced compared to the current imaging. The abdominal CT was unremarkable, and transthoracic ultrasound revealed diffuse hypokinesia in a normal, nondilated left ventricle, with a mild reduction in global systolic function. A subsequent bronchoscopy with bronchoalveolar lavage (BAL) showed no evidence of opportunistic infection,

eosinophilia, alveolar hemorrhage, or proteinaceous material typically seen in pulmonary alveolar proteinosis (PAP).

CASE SUMMARY

The presented case summarizes the disease course of a 53-year-old woman with a refractory autoinflammatory syndrome with the clinical picture of AOSD on high-dose immune suppression, presenting with fever, progressive dyspnea, and hemodynamic instability. Further diagnostic tests revealed elevated inflammatory laboratory parameters, hyperferritinemia, and diffuse progressive interstitial lung disease (ILD) on imaging.

DIFFERENTIAL DIAGNOSIS

On imaging, the interstitial lung abnormalities were predominantly distributed in the mid and lower lung zones in a peribronchovascular pattern, displaying less involvement of the peripheral regions. The primary finding was peribronchovascular ground-glass attenuation, accompanied by areas of crazy paving, characterized by superimposed smooth interlobular septal thickening.

Noteworthy for their absence were diffuse nodular pleural thickening (as seen in sarcoidosis), signs of fibrosis (eg, traction bronchiectasis or honeycombing; as in usual interstitial pneumonia), evident masses or mass-like consolidations (as seen in organizing pneumonia and lymphoma), mosaic attenuation (as in hypersensitivity pneumonitis), and thin-walled cysts (as in lymphocytic interstitial pneumonia). The chronic nature of the lesions made acute alveolar conditions, such as pulmonary edema, acute interstitial pneumonia, or adult respiratory distress syndrome, unlikely differential considerations.[4–6]

Given this context, the differential diagnosis for chronic peribronchovascular ILD was refined to include infection, drug toxicity including eosinophilic pneumonia and AOSD-associated ILD (AOSD-LD) (eg, nonspecific interstitial pneumonia, diffuse pulmonary hemorrhage, lipoid pneumonia, and alveolar proteinosis) among the leading considerations. The following sections provide an in-depth analysis of the remaining differential diagnoses.

Infectious (opportunistic) LD. At initial presentation, the suspicion of infection was high, as the patient had been immunocompromised for an extended period of time and presented with fever and respiratory symptoms. Laboratory investigations demonstrated elevated markers of inflammation and significant leukocytosis. In addition, HRCT revealed diffuse ILD of undetermined etiology. The imaging findings were not consistent with a typical bacterial infection; however, viral or opportunistic pathogens such as cytomegalovirus (CMV) or *Pneumocystis jirovecii* could not be excluded based on imaging alone. All microbiological studies, including blood cultures, BAL, and PCR testing, remained negative. With infection in mind, empirical therapy with amoxicillin/clavulanic acid and azithromycin was initiated and later switched to piperacillin-tazobactam with co-trimoxazole. Additionally, broad-spectrum antifungal and antiviral treatment was added. Despite this treatment the patient deteriorated with progressive hemodynamic instability, increasing parameters of inflammation (CRP >400 mg/L) and increasing oxygen demand. Hence, she was transferred to the ICU for intubation, ventilation, and vasopressin. Overall, the clinical picture did not respond to prolonged treatment with broad-spectrum antibiotics or antiviral/antifungal medication, arguing against an infectious origin of the LD.

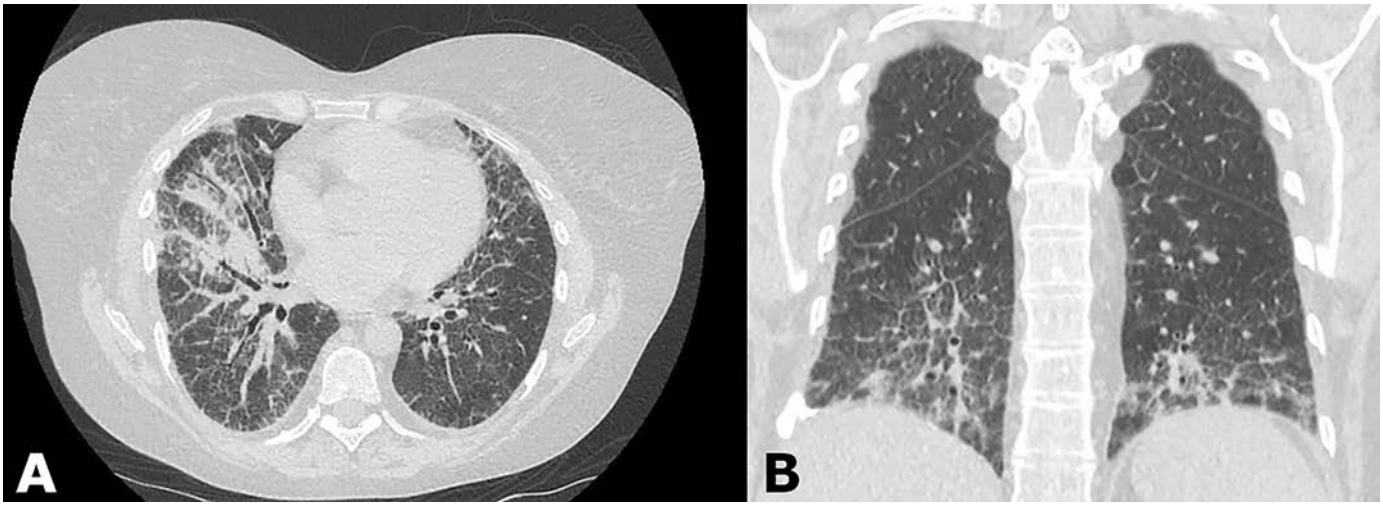


Figure 1. High-resolution computed tomography scan of the chest. (A–B) Chest computed tomography showed diffuse lung disease, predominant in lower and middle lobes with septal thickening involving the periphery of several lobes, some adjacent ground-glass opacities, bilateral peribronchovascular thickening, and some pleural thickening.

Drug-induced ILD or drug reaction with eosinophilia and systemic symptoms. Drug-induced ILD can be suspected when findings consistent with ILD occur after temporal exposure to a culprit drug in absence of other more likely causes, with amelioration upon withdrawal of the suspected culprit. Our patient had been exposed to multiple drugs in the months before admission. Methotrexate, tocilizumab, and anakinra had already been discontinued several months before the event, whereas the ongoing treatment with canakinumab and systemic glucocorticoids was considered less notorious in this respect.[7] Clinical status worsened following the withdrawal of canakinumab.

Similarly, the presence of a drug reaction with eosinophilia and systemic symptoms (DRESS) syndrome, which is considered a delayed T cell–mediated adverse drug reaction, had been taken into consideration. Symptoms of the latter generally consist of fever, rash, facial edema, eosinophilia, and organ involvement, including interstitial pneumonia.[8] Nevertheless, our patient did not display any peripheral eosinophilia or skin rash or acute clubbing at the moment of respiratory distress. Moreover, based on the RegiSCAR score—a scoring system developed by the European Registry of Severe Cutaneous Adverse Reactions (RegiSCAR) to classify DRESS as definite, probable, or not a case according to key diagnostic criteria—our patient's presentation was not suggestive of DRESS.

Notably, experts in the field have proposed that DRESS may contribute to the development of LD in systemic onset juvenile idiopathic arthritis (soJIA), considered the pediatric counterpart of AOSD. These conditions are generally regarded as part of the same spectrum of disease at different ages, suggesting a potential overlap in the pathogenesis of LD.[9,10] Compared with the population of patients with soJIA without LD, children who developed LD had more frequent adverse reactions to biologic therapy, in particular tocilizumab, to which our patient had previously been exposed.[11–13] Half of these reactions fitted the DRESS criteria and were significantly correlated with HLA-DRB1*15.[14,15]

AOSD-LD. AOSD is considered part of the autoinflammatory diseases, a heterogeneous group of disorders characterized by dysregulation of the innate immune system, resulting in recurrent or persistent episodes of seemingly spontaneous systemic inflammation.[16] Other examples hereof are soJIA and TRAPS.[9,13,17,18]

Whereas AOSD is characterized by fever, typical rash, arthritis/arthralgia, leukocytosis, and sore throat, soJIA is characterized by similar systemic features at earlier onset, alongside with lymphadenopathy, hepatosplenomegaly, and/or splenomegaly and serositis.[19,20] The term AOSD refers to a clinical syndrome, as specific genetic mutations have not been identified, unlike in TRAPS. At diagnosis our patient exhibited clear symptoms of autoinflammation, fulfilling the Yamaguchi classification criteria for AOSD.[3,20] One might argue whether a diagnosis of TRAPS needed to be considered. Indeed, the patient exhibited myalgia and a migratory rash, fulfilling the Eurofever/Paediatric Rheumatology International Trials Organisation classification criteria for TRAPS in the absence of a confirmatory TNFRSF1A genotype, with an episode duration exceeding seven days.[1,21] Therefore, our patient also met the classification criteria for TRAPS. Nevertheless, classification criteria are not regarded as diagnostic. Hence, given the clinical phenotype of AOSD and the nonconfirmatory genotype of the mutation, we were more inclined to diagnose AOSD. Moreover, in TRAPS, only anecdotal reports on recurrent pneumonia or persistent cough of unknown origin have been reported.[22]

Lung involvement has been described in 5% to 12% of patients with AOSD, including a range of manifestations such as pleuritis, pulmonary infiltrates, bronchiolitis, and ILD (including endogenous lipoid pneumonia [ELP] and PAP) and pulmonary arterial hypertension.[13,23] Although PAP and ELP—rare pulmonary diseases characterized by the accumulation of proteinaceous (surfactant components) and fatty (cholesterol) material in the alveoli—have been infrequently reported in adults with autoinflammatory diseases, larger cohorts examining lung involvement have been described in the pediatric population with soJIA, in which PAP/ELP were identified as the most common pathology among biopsied cases (Supplementary Table S1).[24–26]

Symptoms of AOSD-LD can include cough, shortness of breath, and pleuritic pain, whereas clinical examination can reveal cyanosis, bibasal crackles, and somewhat more specific digital clubbing.[13,23] Although our patient presented with shortness of breath and bilateral basal fine crackles on auscultation, she did not exhibit cough, pleuritic pain, or clubbing. Laboratory findings, including elevated markers of inflammation, ferritin, IL-6, IL-18, and liver enzymes, as seen in our patient, were also observed in ELP.[13] In case of AOSD-LD, HRCT of the chest can be heterogeneous revealing multilobar, predominantly peripheral septal thickening with or without adjacent ground-glass opacities, crazy paving pattern, peripheral consolidations, peribronchovascular consolidation, and predominantly ground-glass opacities, most frequently seen in the lower lobes.[13,23] Indeed, HRCT of our patient showed diffuse LD in the lower and middle lobes with septal thickening involving the periphery of several lobes, some adjacent ground-glass opacities, bilateral peribronchovascular thickening, and some pleural thickening, fitting with the HRCT patterns described in literature on soJIA/AOSD-LD. On top of that, lung function testing typically shows a restrictive pattern with impaired diffusion capacity for carbon monoxide (DL_{CO}) in AOSD-LD, as observed in the follow-up of this case.[27] BAL fluid did not exhibit the milky appearance suggestive for PAP, nor did it exhibit periodic acid–Schiff-positive reactions, which would have been indicative of extracellular proteinaceous material. However, the absence of this finding should be interpretated with caution, because in a case series of patients with soJIA-LD, BAL fluid rarely contained proteinaceous material and had less lipid-laden macrophages compared to patients with primary PAP.[12]

Additional evaluation. Given the unclear etiology of the LD in our patient and a progressively worsening clinical course, a lung biopsy was performed. Macroscopically, a yellow parenchymal consolidation was noted. Microscopically, organizing diffuse alveolar damage with numerous intra-alveolar cholesterol crystals and, focally, foamy macrophages was observed, consistent with ELP (Figure 2). (Immuno)histochemical examination was negative for fungi, yeasts (including *P jiroveci*), and CMV.

It is worth noting that, although the clinical presentation and radiographic findings in this patient were consistent with soJIA/AOSD-LD, the diagnosis of ELP could only be definitively confirmed through histopathological analysis.[28] Informed consent for publication of this information was obtained from the patient.

DISCUSSION

We describe one of the first cases of ELP in an adult patient with the clinical picture of AOSD and a low penetrance mutation in the TNFRSF1A gene, as seen in TRAPS. A 53-year-old immunocompromised woman with treatment refractory autoinflammatory disease and history of MAS presented to the emergency room with progressive dyspnea and fever. On evaluation, chest-CT showed diffuse LD. An extensive workup, including bronchoscopy with BAL and comprehensive sampling, yielded negative results for an infectious origin. At a later stage, lung biopsy revealed an ELP with intra-alveolar accumulation of cholesterol crystals and foamy macrophages. The patient was then treated with pulse doses of systemic glucocorticoids, followed by induction with cyclophosphamide, maintained on tacrolimus. Upon treatment, our patient improved significantly but retained residual ILD. This case highlights the complex pathophysiology of lung involvement in autoinflammatory diseases, operating at the interface of the innate and adaptive immune system.[29]

The aforementioned inflammatory diseases can be complicated by life-threatening MAS, a hyperinflammatory condition orchestrated by macrophages and T-lymphocytes, resulting in a cytokine storm. MAS typically includes extremely high ferritin levels and/or pancytopenia as most prominent laboratory features and can more specifically show elevation of soluble CD25, whereas natural killer cell activity can be low or absent.[18] MAS has been reported in up to 15% of patients with AOSD and is rarely described in cases of pediatric TRAPS.[30] Strikingly, in soJIA, MAS can even occur in up to 40% of cases.[31] Histopathological findings display hemophagocytic macrophages in bone marrow, lymph nodes, and spleen, all of which are considered a prelude of multiple organ failure with high mortality rates in up to of 8% to 17% in soJIA, increasing up to 10% to 41% in adults.[32] Our patient exhibited several episodes of MAS with corresponding clinical and laboratory manifestations, as well as a suitable cytokine profile throughout her inflammatory flares.

Although isolated cases of nonspecified ILD in AOSD have been described before, the term AOSD-LD has only recently been introduced in parallel to the increased awareness of lung involvement in soJIA.[13,33,34] Recent reports have gone as far as claiming similar prevalence of parenchymal lung involvement in both AOSD and soJIA.[9,35] Interestingly, a retrospective case study in soJIA-LD, focusing on parenchymal lung pathology, predominantly identified PAP and/or ELP as the final diagnosis.[11] In literature on AOSD-LD, to our knowledge, only a few biopsied cases have been reported, of which one case was of ELP and one case was of PAP, and no cases have been described in adult TRAPS.[13,23–25,36] Meanwhile, the other AOSD-LD biopsy specimens have shown nonspecific findings of interstitial fibrosis and chronic inflammation.[23,33,34,37–40]

As previously mentioned, ELP is a rare form of ILD caused by lipid accumulation within the lung. This can be observed, for example, in airway obstruction or systemic diseases with chronic

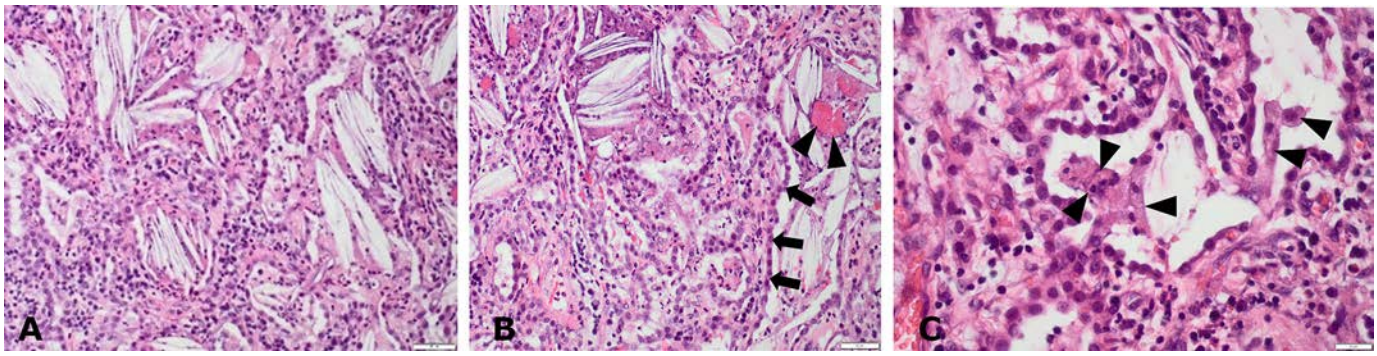


Figure 2. Histopathology of the lung biopsy. (A) Deposition of needle-shaped cholesterol crystals in the alveolar spaces. The interstitium of the alveoli is inflamed, showing mononuclear inflammatory cells and some neutrophils. (B) Area in which both organizing diffuse alveolar damage and deposition of cholesterol crystals are seen. The alveolar walls are thickened and there is obvious type II pneumocyte hyperplasia (arrows). Some residual hyaline material is indicated by arrowheads. (C) At high magnification, clusters of foamy macrophages are visualized between the cholesterol crystals (arrowheads).

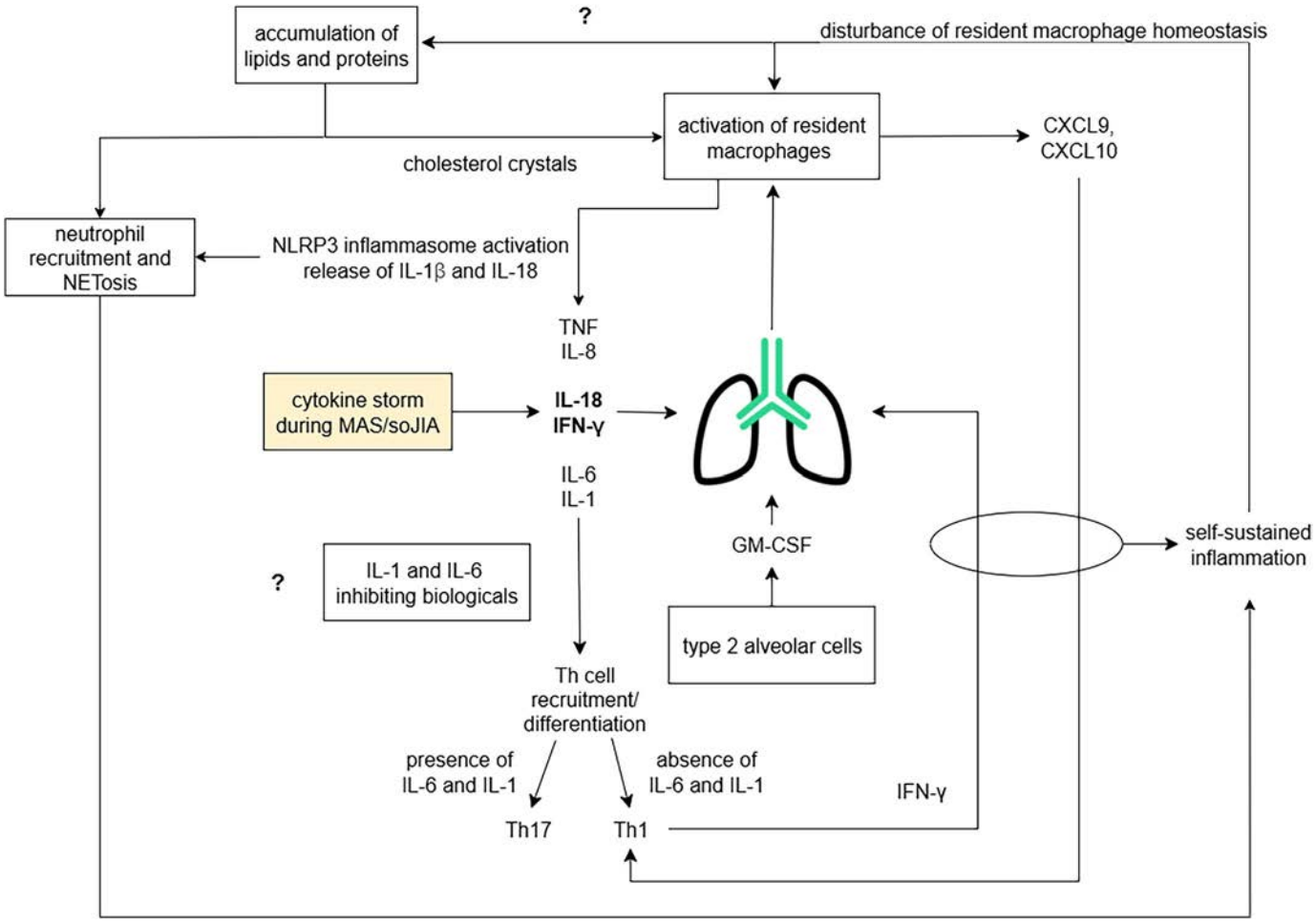


Figure 3. Possible mechanism of soJIA-LD. The cytokine storm in MAS attributes an important role to IL-18 and IFN-γ, which activate alveolar macrophages and other immune cells in the lungs. Presence of IL-6 and IL-1 favors differentiation into Th17 cells, whereas absence promotes polarization toward Th1 cells.[45] The latter produce IFN-γ, further enhancing macrophage activation. These activated immune cells release cytokines and chemokines, including CXCL9 and CXCL10. This recruits even more Th1-lymphocytes, leading to a self-sustained inflammatory response in the lung. This proinflammatory environment disturbs homeostasis of resident alveolar macrophages, resulting in accumulation of lipids and proteins, leading to pulmonary alveolar proteinosis–like features.[12] These lipids can induce NLRP3 inflammasome formation in macrophages and neutrophils, respectively, leading to pyroptosis and NETosis, further sustaining the inflammation.[48–50] GM-CSF, granulocyte-macrophage colony-stimulating factor; IFN-γ, interferon gamma; IL, interleukin; LD, lung disease; MAS, macrophage activation syndrome; NET, neutrophil extracellular traps; NLRP3, NOD-like receptor family pyrin domain containing 3; soJIA, systemic-onset juvenile idiopathic arthritis; TNF, tumor necrosis factor. Figure by Katrien Slabbynck, adapted from Schulert et al.[12]

inflammation among other causes. These endogenous lipids consist of the body's own cholesterol, originating from type II epithelial cells within the alveolar walls. Furthermore, intra-alveolar macrophages engulf the accumulated lipids in the alveolar cavities forming lipid-laden macrophages or so-called foam cells. There is no established standard of care for the treatment of ELP, although the first step is to address the underlying cause. Different strategies have been explored, including glucocorticoids, immunoglobulins, whole-lung lavage, lung transplantation, and even anti–tumor necrosis factor.[27] Our patient has been successfully treated with high-dose glucocorticoids and cyclophosphamide induction followed by tacrolimus maintenance therapy. As demonstrated in our patient, pulmonary manifestations of AOSD can be severe and even lethal, therefore, in concordance with soJIA-LD, screening should be considered to facilitate early detection and intervention to prevent long-term morbidity and mortality.[13,41]

The precise mechanism of lung disease in autoinflammatory conditions remains unclear and poorly characterized, but there are several things we can learn about possible mechanisms from the literature on soJIA. First, multiple studies have shown a link between soJIA-LD and disease severity. Compared to the population of patients with soJIA without LD, children who developed LD were younger at diagnosis and had a more active disease, and MAS occurred more frequently.[11–13,15,42–44] Concordantly, our patient had a treatment refractory disease with the occurrence of several flares of MAS, an utterance of severe

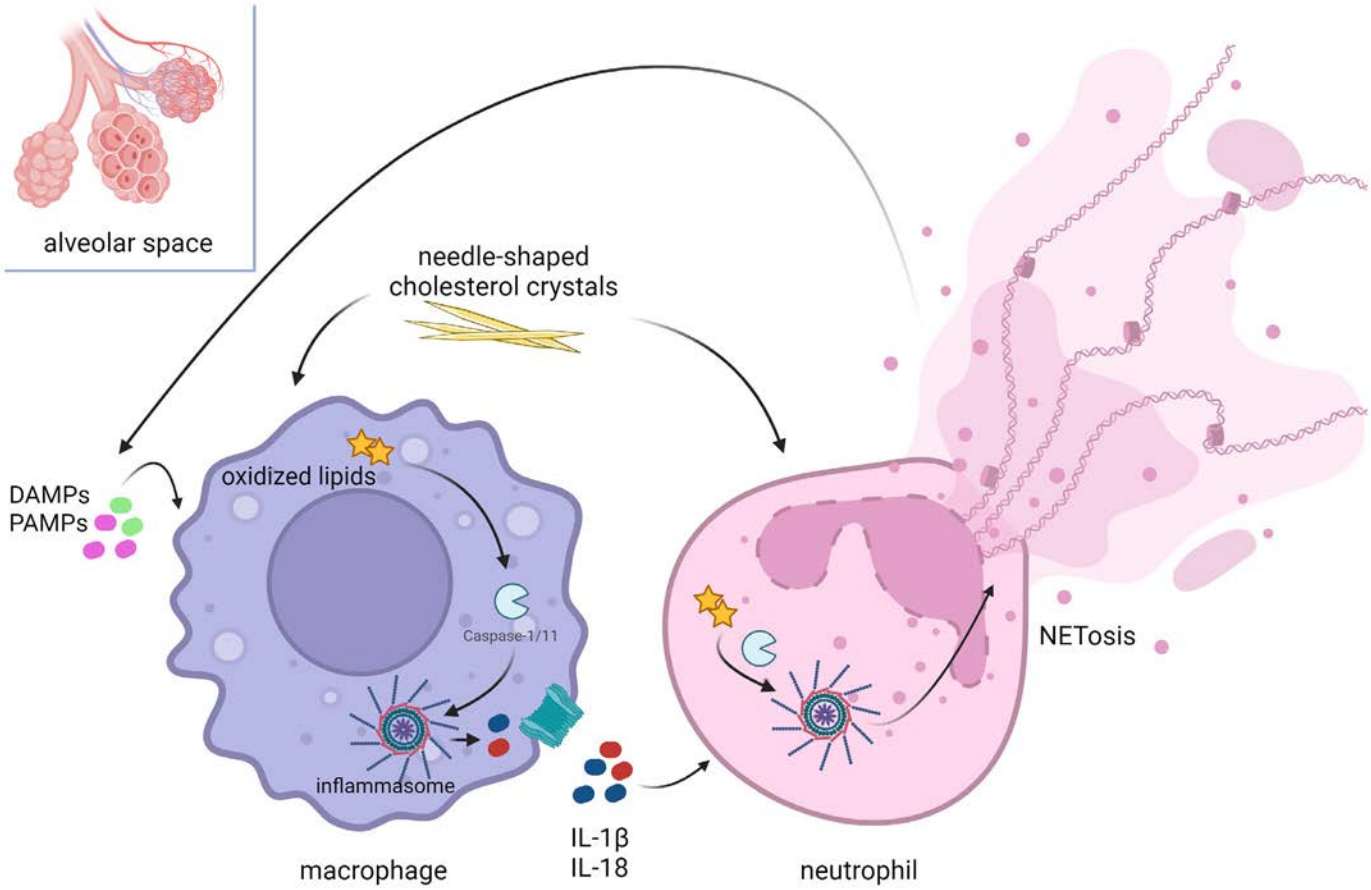


Figure 4. Possible mechanisms of NLRP3 inflammasome activation by cholesterol crystals in macrophages and neutrophils in the alveolar space and their crosstalk. Cholesterol crystals in the alveolar space are engulfed by macrophages and neutrophils. The oxidized lipids lead to inflammasome activation and caspase-dependent cleavage of prointerleukins into their active form as well as pore formation leading to pyroptosis, NETosis, and release of IL-18 and IL-1β. The expulsion of intracellular material again leads to DAMPs and further activation of inflammation.[51] DAMP, damage-associated molecular pattern; IL, interleukin; NET, neutrophil extracellular traps; PAMP, pathogen-associated molecular pattern. Figure by Katrien Slabbynck with BioRender, adapted from Tall et al.[51] Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25570/abstract.

disease. The cytokine storm occurring in MAS may lead to an altered cytokine environment, dominated by interferon (IFN)-γ and IL-18.[12] In the proximity of specific cytokines, $CD4^+$ T cells are known to differentiate into certain subtypes. This mechanism may alter the phenotypes of immune cells in the lung, contributing to ELP through an as-yet unidentified pathway.[45] Supportive of this theory, some studies showed that higher serum IL-18, elevated ferritin, and lymphopenia preceded the diagnosis of LD.[11,12] Accordingly, the cytokine analysis in our patient showed strong elevation of IL-1RA, IL-18, IL-6 and CXCL9, an IFN-dependent chemokine, as seen in a cytokine storm such as MAS. Based on these data, IL-18 blocking agents may have a role in patients with concomitant LD. Indeed, a recent article reported a case of soJIA-LD that showed a favorable response to combined IL-18 and IL-1β blockade.[46]

Second, it has been suggested in some studies that soJIA-LD might be related to the coincident increased use of biologics. Saper et al reported that 80% of lung biopsies in patients with soJIA-LD with pre-exposure to cytokine blockers showed PAP/ELP features compared to only 36% in patients with cytokine-blocking treatment-naive soJIA-LD.[11] The DRESS hypothesis, as previously mentioned in the Differential Diagnosis section, was proposed to explain this link. Nevertheless, LD was also seen in patients who had not been exposed to cytokine-blocking agents, arguing against their causative role.[42,44,47] Certainly, the use of biologics is more likely to reflect the severity of the disease, necessitating more advanced therapeutic interventions.

As presented, our patient had been treated with several biologic treatments including tocilizumab, anakinra, and canakinumab before being diagnosed with ELP. Nevertheless, before the diagnosis of AOSD and before the initiation of any biologic treatment, chest-CT revealed bilateral septal thickening in the apical lobes, already suggesting the presence of AOSD-LD early in the

disease course. In retrospect, the initial presentation with suspected *C psittacosis* infection might already have been an underappreciated manifestation of AOSD-LD. A graphical overview of possible mechanisms is shown in Figure 3.

Third, there might be a role for the cholesterol crystals in the alveolar cavities in itself as a self-maintaining trigger for chronic inflammation. It is shown that cholesterol crystals can lead to NLRP3 inflammasome activation with subsequent release of cytokines IL-1β and IL-18 and pyroptosis, an inflammatory form of cell death.[48,49] A possible mechanism of NLRP3 inflammasome activation is shown in Figure 4. Our patient's lung biopsy showed marked neutrophilic infiltration, with serum cytokine panels displaying elevated IL-18. Through a similar mechanism, the cholesterol crystals might induce neutrophil extracellular traps-(NET)osis of those neutrophils leading to an exacerbation of proinflammatory responses.[50]

In summary, we describe a patient with refractory autoinflammatory syndrome with the clinical picture of AOSD complicated by several episodes of MAS, presenting with respiratory distress secondary to ELP. This case illustrates the disease continuum of autoinflammation in both adults and children and highlights the complex and unidentified pathophysiology of LD herein. Importantly, disease severity and/or MAS seem to be a key precipitating factor in predisposed patients. Overall, more research is needed to disentangle the complex interplay between host and environmental factors on the pathophysiology of ELP across the disease spectrum. A role may be assigned to interleukin-blocking agents, although their role as proxy for severe disease or as a true catalysator needs to be elucidated.

THE PATIENT'S COURSE

The patient was pulsed with 1,000 mg methylprednisolone intravenously (IV) for three days, with considerable response. Because of the life-threatening nature of her disease and refractory status despite exhaustive anti-inflammatory treatment against the underlying autoinflammatory syndrome, she underwent induction therapy with cyclophosphamide IV 500 mg/m^2 body surface area every four weeks for four cycles with subsequent tacrolimus maintenance therapy. This regimen was empirically employed based on experience in other systemic inflammatory diseases. Arguably, this approach was effective, although the EUROLUPUS regimen may have been a suitable alternative. Nevertheless, due to relapse of her inflammatory symptoms a couple of weeks after completion of the cyclophosphamide induction, canakinumab was reintroduced. Since then, the patient has remained out of hospital, with successful tapering of the glucocorticoids. During follow-up, lung function examination revealed restrictive impairment with a forced vital capacity of 45%, a total lung capacity of 64%, and significant diminished DL_{CO} of 55%. Clinically, radiologically, and functionally significant residual ILD persists, although features of fibrosis, such as traction bronchiectasis and honeycombing, remain absent.

FINAL DIAGNOSIS

Endogenous lipoid pneumonia in adult autoinflammatory disease with the clinical picture of adult-onset Still disease.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Varkas 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 Helsinki of requirements.

REFERENCES

1. Cudrici C, Deuitch N, Aksentijevich I. Revisiting TNF receptor-associated periodic syndrome (TRAPS): current perspectives. Int J Mol Sci 2020;21(9):3263.
2. Gaggiano C, Vitale A, Obici L, et al. Clinical features at onset and genetic characterization of pediatric and adult patients with TNF-α receptor-associated periodic syndrome (TRAPS): a series of 80 cases from the aida network. Mediators Inflamm 2020;2020:8562485.
3. Lachmann HJ, Papa R, Gerhold K, et al; Paediatric Rheumatology International Trials Organisation (PRINTO), the EUROTRAPS and the Eurofever Project. The phenotype of TNF receptor-associated autoinflammatory syndrome (TRAPS) at presentation: a series of 158 cases from the Eurofever/EUROTRAPS international registry. Ann Rheum Dis 2014;73(12):2160–2167.
4. Le L, Narula N, Zhou F, et al. Diseases involving the lung peribronchovascular region: a CT imaging pathologic classification. Chest 2024;166(4):802–820.
5. Cozzi D, Cavigli E, Moroni C, et al. Ground-glass opacity (GGO): a review of the differential diagnosis in the era of COVID-19. Jpn J Radiol 2021;39(8):721–732.
6. Ko JP, Girvin F, Moore W, et al. Approach to peribronchovascular disease on CT. Semin Ultrasound CT MR 2019;40(3):187–199.
7. Skeoch S, Weatherley N, Swift AJ, et al. Drug-induced interstitial lung disease: a systematic review. J Clin Med 2018;7(10):356.
8. Hama N, Abe R, Gibson A, et al. Drug-induced hypersensitivity syndrome (DIHS)/drug reaction with eosinophilia and systemic symptoms (DRESS): clinical features and pathogenesis. J Allergy Clin Immunol Pract 2022;10(5):1155–1167.e5.
9. De Matteis A, Bindoli S, De Benedetti F, et al. Systemic juvenile idiopathic arthritis and adult-onset Still's disease are the same disease: evidence from systematic reviews and meta-analyses informing the 2023 EULAR/PReS recommendations for the diagnosis and management of Still's disease. Ann Rheum Dis 2024;83(12):1748–1761.
10. Fautrel B, Mitrovic S, De Matteis A, et al. EULAR/PReS recommendations for the diagnosis and management of Still's disease, comprising systemic juvenile idiopathic arthritis and adult-onset Still's disease. Ann Rheum Dis 2024;83(12):1614–1627.
11. Saper VE, Chen G, Deutsch GH, et al; Childhood Arthritis and Rheumatology Research Alliance Registry Investigators. Emergent high

fatality lung disease in systemic juvenile arthritis. Ann Rheum Dis 2019;78(12):1722–1731.

12. Schulert GS, Yasin S, Carey B, et al. Systemic juvenile idiopathic arthritis-associated lung disease: characterization and risk factors. Arthritis Rheumatol 2019;71(11):1943–1954.
13. Nies JF, Schneider U, Krusche M. Rare, rarer, lung involvement in adult-onset Still's disease: a mini-review. Front Med (Lausanne) 2022;9:989777.
14. Petrongari D, Di Filippo P, Misticoni F, et al. Lung involvement in systemic juvenile idiopathic arthritis: a narrative review. Diagnostics (Basel) 2022;12(12):3095.
15. Lerman AM, Mahmud SA, Alfath Z, et al. HLA-DRB1*15 and eosinophilia are common among patients with systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2023;75(10):2082–2087.
16. Betrains A, Staels F, Schrijvers R, et al. Systemic autoinflammatory disease in adults. Autoimmun Rev 2021;20(4):102774.
17. Vitale A, Caggiano V, Lopalco G, et al. Still's disease continuum from childhood to elderly: data from the international AIDA Network Still's disease registry. RMD Open 2023;9(4):e003578.
18. Ravelli A, Minoia F, Davì S, et al; Paediatric Rheumatology International Trials Organisation; Childhood Arthritis and Rheumatology Research Alliance; Pediatric Rheumatology Collaborative Study Group; Histiocyte Society. 2016 classification criteria for macrophage activation syndrome complicating systemic juvenile idiopathic arthritis: a European League Against Rheumatism/American College of Rheumatology/Paediatric Rheumatology International Trials Organisation collaborative initiative. Arthritis Rheumatol 2016;68(3):566–576.
19. Lee JJY, Schneider R. Systemic juvenile idiopathic arthritis. Pediatr Clin North Am 2018;65(4):691–709.
20. Yamaguchi M, Ohta A, Tsunematsu T, et al. Preliminary criteria for classification of adult Still's disease. J Rheumatol 1992;19(3):424–430.
21. Gattorno M, Hofer M, Federici S, et al; Eurofever Registry and the Paediatric Rheumatology International Trials Organisation (PRINTO). Classification criteria for autoinflammatory recurrent fevers. Ann Rheum Dis 2019;78(8):1025–1032.
22. Deuitch N, Cudrici C, Ombrello A, et al. TNF receptor-associated periodic fever syndrome. In: Adam MP, Everman DB, Mirzaa GM, et al, eds. GeneReviews. University of Washington; 1993–2025.
23. Gerfaud-Valentin M, Cottin V, Jamilloux Y, et al. Parenchymal lung involvement in adult-onset Still disease: a STROBE-compliant case series and literature review. Medicine (Baltimore) 2016;95(30):e4258.
24. Sami R, Zohal M, Mohammadi F. Adult onset still disease associated with endogenous lipoid pneumonia. Adv Respir Med 2019;87(1):50–53.
25. Bindoli S, Lococo S, Calabrese F, et al. Pulmonary alveolar proteinosis in an adult patient affected by Still's disease and recurrent episodes of macrophage activation syndrome. Joint Bone Spine 2024;91(1):105654.
26. Jouneau S, Ménard C, Lederlin M. Pulmonary alveolar proteinosis. Respirology 2020;25(8):816–826.
27. Hadda V, Khilnani GC. Lipoid pneumonia: an overview. Expert Rev Respir Med 2010;4(6):799–807.
28. Beck LR, Landsberg D. Lipoid pneumonia. In: Statpearls. StatPearls Publishing LLC; 2022.
29. Peckham D, Scambler T, Savic S, et al. The burgeoning field of innate immune-mediated disease and autoinflammation. J Pathol 2017;241(2):123–139.
30. Horneff G, Rhouma A, Weber C, et al. Macrophage activation syndrome as the initial manifestation of tumour necrosis factor receptor 1-associated periodic syndrome (TRAPS). Clin Exp Rheumatol 2013;31(3 suppl 77):99–102.
31. Ruscitti P, Rago C, Breda L, et al. Macrophage activation syndrome in Still's disease: analysis of clinical characteristics and survival in paediatric and adult patients. Clin Rheumatol 2017;36(12):2839–2845.
32. Minoia F, Davì S, Horne A, et al; Pediatric Rheumatology International Trials Organization; Childhood Arthritis and Rheumatology Research Alliance; Pediatric Rheumatology Collaborative Study Group; Histiocyte Society. Clinical features, treatment, and outcome of macrophage activation syndrome complicating systemic juvenile idiopathic arthritis: a multinational, multicenter study of 362 patients. Arthritis Rheumatol 2014;66(11):3160–3169.
33. Corbett AJ, Zizic TM, Stevens MB. Adult-onset Still's disease with an associated severe restrictive pulmonary defect: a case report. Ann Rheum Dis 1983;42(4):452–454.
34. Van Hoeyweghen RJ, De Clerck LS, Van Offel JF, et al. Interstitial lung disease and adult-onset Still's disease. Clin Rheumatol 1993;12(3):418–421.
35. Ruscitti P, Natoli V, Consolaro A, et al. Disparities in the prevalence of clinical features between systemic juvenile idiopathic arthritis and adult-onset Still's disease. Rheumatology (Oxford) 2022;61(10):4124–4129.
36. Ito Y, Nakahara H, Nakajima A. Development of pulmonary alveolar proteinosis in a patient with adult-onset Still disease treated with tocilizumab. J Rheumatol 2023;50(2):288–289.
37. Martín-Nares E, López-Iñiguez A, Castro-Gallegos PE, et al. Adult-onset Still disease presenting with insterstitial lung disease and dermatomyositis-like atypical persistent eruption successfully treated with mycophenolate mofetil. J Clin Rheumatol 2019;25(8):e136–e140.
38. Suyama T, Yokoyama K. [Interstitial pneumonitis detected by bronchoalveolar lavage and transbronchial lung biopsy in adult-onset Still's disease]. Nihon Kokyuki Gakkai Zasshi 1998;36(6):545–550.
39. Takakuwa Y, Hanaoka H, Kiyokawa T, et al. Adult-onset still's disease-associated interstitial lung disease represents severe phenotype of the disease with higher rate of haemophagocytic syndrome and relapse. Clin Exp Rheumatol. 2019;37 Suppl 121(6):23–27.
40. Hijikata N, Takayanagi N, Sugita Y, et al. Adult-onset still's disease with pulmonary involvement. J Bronchology Interv Pulmonol 2009;16(4):277–282.
41. Wobma H, Bachrach R, Farrell J, et al. Development of a screening algorithm for lung disease in systemic juvenile idiopathic arthritis. ACR Open Rheumatol 2023;5(10):556–562.
42. Kimura Y, Weiss JE, Haroldson KL, et al; Childhood Arthritis Rheumatology Research Alliance Carra Net Investigators. Pulmonary hypertension and other potentially fatal pulmonary complications in systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2013;65(5):745–752.
43. Wobma H, Arvila SR, Taylor ML, et al. Incidence and risk factors for eosinophilia and lung disease in biologic-exposed children with systemic juvenile idiopathic arthritis. Arthritis Care Res (Hoboken) 2023;75(10):2063–2072.
44. Zhan W, Yang J, Qiu L, et al. Clinical characteristics and prognosis of interstitial lung disease in systemic juvenile idiopathic arthritis: a two-center retrospective observational cohort study. Pediatr Rheumatol Online J 2024;22(1):96.
45. Binstadt BA, Nigrovic PA. The conundrum of lung disease and drug hypersensitivity-like reactions in systemic juvenile idiopathic arthritis. Arthritis Rheumatol 2022;74(7):1122–1131.
46. Rood JE, Rezk A, Pogoriler J, et al. Improvement of refractory systemic juvenile idiopathic arthritis-associated lung disease with single-agent blockade of IL-1β and IL-18. J Clin Immunol 2023;43(1):101–108.
47. Huang Y, Sompii-Montgomery L, Patti J, et al. Disease course, treatments, and outcomes of children with systemic juvenile idiopathic

arthritis-associated lung disease. Arthritis Care Res (Hoboken) 2024; 76(3):328–339.

48. Duewell P, Kono H, Rayner KJ, et al. NLRP3 inflammasomes are required for atherogenesis and activated by cholesterol crystals. Nature 2010;464(7293):1357–1361.
49. Rajamäki K, Lappalainen J, Oörni K, et al. Cholesterol crystals activate the NLRP3 inflammasome in human macrophages: a novel link between cholesterol metabolism and inflammation. PLoS One 2010; 5(7):e11765.
50. Münzer P, Negro R, Fukui S, et al. NLRP3 inflammasome assembly in neutrophils is supported by PAD4 and promotes NETosis under sterile conditions. Front Immunol 2021;12:683803.
51. Tall AR, Westerterp M. Inflammasomes, neutrophil extracellular traps, and cholesterol. J Lipid Res 2019;60(4):721–727.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1194–1203
DOI 10.1002/acr.25555

Real-World Cost-Effectiveness of a Standardized Education and Exercise Therapy Program for Hip and Knee Osteoarthritis Compared to Usual Care

Darren R. Mazzei,[1] Jackie L. Whittaker,[2] Peter Faris,[3] Tracy Wasylak,[3] and Deborah A. Marshall[1]

Objective. We estimated the real-world cost-effectiveness of a standardized education and exercise therapy program (GLA:D) compared to usual care (UC) for people managing hip and/or knee osteoarthritis (HKOA).

Methods. We used a prospective matched cohort design to recruit people (aged >45 years) diagnosed with HKOA who used GLA:D or UC (not on a surgical waitlist) throughout Alberta, Canada. Demographics, pain, function, quality of life, and an HKOA-related cost questionnaire were administered over 12 months. The primary Ministry of Health (MOH) perspective used administrative data to estimate all public health care costs. The secondary health care perspective included MOH, private insurance, and out-of-pocket costs. We calculated our cost-effectiveness measure, incremental net monetary benefit (INMB) in Canadian dollars, over 12 months with a $30,000/quality-adjusted life years (QALY) willingness to pay threshold and adjusted for the differences between cohorts. A Markov model was used to extend INMB over a lifetime time horizon (3% discounting). Model uncertainty was explored by probabilistic sensitivity analyses.

Results. A total of 254 participants (GLA:D n = 127, UC n = 127; 72% female), with a mean age of 64.3 years (95% confidence interval [CI]: 63.1–65.5), diagnosed with knee osteoarthritis (63%), hip osteoarthritis (24%), or both (13%) were observed for a mean of 5.5 years (95% CI: 4.8–6.3). The adjusted INMB of GLA:D compared to UC was $6,065 (95% CI: $3,648–$8,482) and $499 (95% CI: –$2,913 to $3,912) from an MOH and health care perspective over 12 months and $6,574 and $1,775 over a lifetime with 54% and 51% probability of being cost-effective using a threshold of willingness to pay of $30,000 per QALY.

Conclusions. GLA:D had a positive INMB compared to UC from the MOH perspective over 12 months. The INMB remained positive but was less certain over a lifetime or when out-of-pocket and private insurance costs were considered.

INTRODUCTION

Osteoarthritis (OA) is highly prevalent, placing a significant burden on people and health systems.[1,2] Clinical guidelines recommend education and exercise therapy as first-line treatments for everyone with hip and knee OA.[3] Attempting first-line treatments is also an eligibility criterion for appropriate total joint replacement (TJR),[4] but 40% of Canadians with knee OA have not attempted first-line treatments before having surgery.[5] These findings are not unique; first-line treatments are underused globally.[6] Barriers to first-line treatment include knowledge gaps, expectations, referral patterns, availability, and costs.[7,8] Increasing uptake by reducing barriers is a global priority.[9]

Integrating standardized education and exercise therapy programs into funded or insured clinical pathways can increase access to first-line treatments. Most economic evaluations

Dr Mazzei's work was supported by the Arthritis Society Training Graduate PhD Salary Award, Cumming School of Medicine Graduate Student Scholarship, Arthur J. E. Child Chair Rheumatology Outcomes Research, and Canadian Institutes of Health Research. Dr Whittaker's work was supported by the Michael Smith Health Research BC and the Arthritis Society. Dr Marshall's work was supported by the Arthur J. E. Child Chair in Rheumatology and a Canada Research Chair in Health Systems and Services Research.

[1]Darren R. Mazzei, PhD, Deborah A. Marshall, PhD: University of Calgary, Calgary, Alberta, Canada, and Arthritis Research Canada, Vancouver, British Columbia, Canada; [2]Jackie L. Whittaker, PhD: University of Calgary, Calgary, Alberta, Canada, and Arthritis Research Canada and University of British Columbia, Vancouver, British Columbia, Canada; [3]Peter Faris, PhD, Tracy Wasylak, MSc: University of Calgary and Alberta Health Services, Calgary, Alberta, 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.25555).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25555.

Address correspondence via email to Deborah A. Marshall, PhD, at damarsha@ucalgary.ca.

Submitted for publication June 21, 2024; accepted in revised form March 28, 2025.

SIGNIFICANCE & INNOVATIONS

- Real-world economic evaluation showed GLA:D produced more value for money than usual care in a publicly insured health care system.
- Publicly funding structured education and exercise therapy programs like GLA:D would be an efficient use of health system resources.

alongside randomized controlled trials (RCTs) show that standardized education and exercise therapy programs are cost-effective in different health systems.[10] Previous economic evaluations used RCTs to collect cost and effects, but RCTs may have limited generalizability because they use controlled environments with targeted populations to evaluate efficacy. The comparator and sample population also impact cost-effectiveness results.[10] Danish RCTs showed that a 12-week individualized nonsurgical knee OA intervention, including exercise, education, insoles, dietary advice, and/or pain medication, was cost-effective compared to written advice in people not eligible for TJR[11] but was not cost-effective compared to TJR in an eligible sample.[12] Danish researchers then created an eight-week standardized group program called Good Life with osteoArthritis in Denmark (GLA:D) to implement high-quality hip and knee OA care in the Danish health system.[13] GLA:D consists of two education sessions and 12 supervised neuromuscular exercise sessions delivered twice per week.[13] GLA:D has spread to 10 countries, and 85,000 people have taken the program.[14] Implementing GLA:D presents an opportunity to evaluate the cost-effectiveness of a standardized education and exercise therapy program in the real world. We evaluated the incremental net monetary benefit (INMB) of GLA:D in comparison to usual care (UC) for managing hip and/or knee OA in the community.

METHODS

Study design. We followed the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) for transparently reporting health economics evaluations[15] (Supplementary File 1). We compared real-world data from a cohort of GLA:D participants with hip and/or knee OA to a cohort of people managing hip and/or knee OA with UC in Alberta, Canada. UC was defined as any community-based service people used to manage their OA symptoms before a TJR. Annually, 170,000 people see a family physician for hip and/or knee OA in Alberta, Canada, whereas 500 people participate in GLA:D (0.3% of the eligible population).[16] During the study, GLA:D was offered in-person or virtually at 68 clinics in Alberta, Canada.[17]

Participants. Participants were enrolled in the study if they were ≥45 years of age, were diagnosed with hip and/or knee OA by a health professional (inclusion criteria to participate in GLA:D), had a natural joint as their primary complaint, were not waiting for a TJR (waiting defined as eligible and waiting for TJR date), and were able to read/comprehend English.

Recruitment. Recruitment happened between January 4, 2021, and January 4, 2022. Participants were recruited with posters and in-person by clinicians or the lead author (DRM) at primary care, rehabilitation, and orthopedic surgeon clinics. GLA:D recruitment posters were only at GLA:D sites. UC posters were at clinics, community pharmacies, and recreation centers and posted on the internet. The Alberta government restricted access to group-based programs during the COVID-19 pandemic, which coincided with the recruitment window.

Data collection. Demographics, including date of birth, sex, height, weight, education, employment status, private insurance, comorbidities, physical activity, fear of joint damage, and previous knee surgeries, were collected at baseline. This study does not evaluate gender because the GLA:D Canada database only collects information about sex. A cost questionnaire was developed to collect patient-reported health care use and out-of-pocket costs for physicians, allied health, diagnostic imaging, injection, medications, and medical devices (Supplementary File 2). The cost questionnaire, European Quality of Life 5-Dimension five-level version (EQ-5D-5L), 12-item Hip Injury and Osteoarthritis Outcome Score (HOOS-12), and 12-item Knee Injury and Osteoarthritis Outcome Score (KOOS-12) were collected at baseline and 3, 6, 9, and 12 months.

All surveys were completed electronically. We extracted data from the GLA:D Canada national database to prevent duplicate collection.[18] GLA:D Canada uses DADOS to collect demographics and outcome measures before beginning the program, then at 3 and 12 months after the preprogram survey date for all GLA:D participants. In addition, GLA:D participants used REDCap to complete cost questionnaires at all time points, as well as outcomes at six and nine months. UC participants used REDCap to collect all demographics, outcomes, and cost questionnaires. DADOS is hosted at the University Health Network.[19] REDCap is hosted at the University of Calgary Clinical Research Unit.[20]

All participants provided consent to link self-reported data to administrative data using personal health numbers. Administrative data linkage is used to collect all publicly funded health care resource use to estimate public-payer health care costs. We used four administrative databases: Alberta Healthcare Insurance Plan (AHCIP) Population Registry, Ambulatory Care Classification/National Ambulatory Care Reporting System (NACRS), Discharge Abstract Database (DAD), and Practitioner Claims (PC). The AHCIP Population Registry contains individual-level demographic information, including personal health number, age, sex, and death, for all patients covered by the insurance plan. NACRS includes provider information, dates, diagnosis, and procedure

codes for all ambulatory services and day surgeries in publicly insured facilities. DAD includes dates, provider information, diagnosis, examination, procedures, and discharge information for all inpatient hospital stays. PC includes all fee-for-service and shadow-billed data submitted for public reimbursement by physicians.

Health outcomes. Quality-adjusted life years (QALY) were calculated as the primary health outcome measure. Participants used the EQ-5D-5L to select from five possible responses on five health domains, producing 3,125 possible health states.[21] Health states were weighted with estimates from the general Canadian population.[22,23] An area under the curve calculation was used to estimate QALYs gained over 12 months.

Clinical effectiveness was evaluated by calculating the change in pain, function, quality of life (QoL), and a summary score over 12 months using the HOOS-12 and KOOS-12. Question responses were summed and then divided by the optimal score to produce a score from 0 (worst) to 100 (best).[24]

Costs. Ministry of Health (MOH) and health care perspectives were used in the cost calculation. The MOH perspective includes all publicly funded health care costs. The health care perspective includes MOH, private insurance, and out-of-pocket costs.

MOH was the primary perspective because it is the reference case in Canada.[25] OA services are funded by various public and private providers in Alberta, Canada (population 4.4 million). AHCIP provides 100% publicly funded coverage for medically necessary care to all citizens except Indigenous peoples, members of the military, and people who opt out of coverage. OA services covered by the AHCIP include physician consultations, diagnostic imaging, hospital stays, surgeries, inpatient medication, and inpatient rehabilitation. People older than 65 years receive 70% publicly insured coverage for prescription medications. Some rural hospitals and suburban Primary Care Networks provided limited access to publicly funded outpatient rehabilitation.

MOH costs were estimated from de-identified patient-level OA-related health care use collected by administrative data. Physician claims were costed using the Alberta Health Insurance Plan: Schedule of Medical Benefits.[26] Resource use identified in NACRS and DAD was costed by the case mix grouper method.[27] Allied health visits with $0 out-of-pocket costs were assumed to receive publicly funded care at the average hourly salary of an allied health professional in the public health care system (ie, average physiotherapy wage is $43.48/hour plus 20% for benefits). Costs were calculated in 2022 Canadian dollars, and all values are listed in Canadian dollars.

Out-of-pocket and private insurance costs were estimated from the cost questionnaire. Participants reported the number of visits and out-of-pocket cost per visit for all OA-related services used. The number of visits, unit costs, and private insurance copay were multiplied to estimate the total cost for allied health professional visits. The cost of a one-month supply for over-the-counter, prescription, gastric protection, and sleep/mood medications was reported then multiplied by three and the private insurance copay to estimate three-month medication costs.

Sample size. Economic evaluations are powered to detect a statistical difference in the effect size for the clinical effectiveness measure, and estimation techniques are used to evaluate cost-effectiveness.[28] A sample of 24 pairs is required to detect a 10-point difference (with an SD of 14) on the HOOS-12/KOOS-12 using a paired two-sample test of means powered at 80% with a statistical significance of $\alpha = 0.05$ and assuming a correlation of 0.30. Increasing the sample size by 20% for dropout and rounding up required 30 pairs in four matched categories (women with knee OA, men with knee OA, women with hip OA, and men with OA). We needed a minimum of 120 participants in each cohort but kept recruitment open for 12 months to recruit additional male participants, who were underrepresented in our sample.

Statistical analysis. Data cleaning and analysis were performed in R version 4.2.2 (2022-10-31 ucrt). The study was approved by the Research Ethics Board at the University of Calgary (REB 20-0613).

Missing data and outliers. We visually inspected patterns in missing and complete data. Missingness did not appear different between cohorts, so data were assumed to be missing at random. Missing data were imputed at three-month intervals using multiple imputations. Convergence was assessed by checking summaries of imputed values across iterations. We generated five imputed data sets, and pooled estimates were reported. Total MOH and health care costs greater than 3.5 absolute deviations from the median were removed. Absolute deviation from the median allowed us to retain the sample size because it is a more robust measure of dispersion than SDs from the mean.[29]

Cost-effectiveness analysis over 12 months. INMB was calculated as the primary cost-effectiveness measure (Supplementary Figure 1). Net benefit converts the cost-effectiveness ratio into a linear expression, which enables the use of regression methods, adjustment for differences between cohorts, and subgroup analysis.[30] The INMB is also easier to interpret because dollar values above $0 mean the intervention is more efficient than the comparator.[30] We used $30,000/QALY as the decision-maker's willingness to pay (WTP) threshold,[31,32] which is the maximum amount a decision-maker will pay for additional health benefits.[30] Linear regression was used to explore

how baseline health utilities, affected joint (hip, knee or hip and knee), and sex (male or female) impacted cost-effectiveness results. Assumptions for linear regression were achieved. Scenario analysis compared participants with complete and imputed data with or without proceeding to TJR in both perspectives. Discounting was not applied over 12 months.

Cost-effectiveness analysis over lifetime. Due to the chronic nature of OA, it is recommended to extend empirical findings past the observation period to estimate how the intervention impacts health and costs to end of life.[33] A Markov model was built using the heemod package in R to extend cost and health outcomes to the lifetime time horizon.[34] People with OA are a heterogeneous population who exhibit high variability in disease progression, severity, time since diagnosis, and response to treatment.[35] We reduced model uncertainty by modeling the health services people use to manage OA instead of modeling disease progression.[30] The model consists of four health states: community management, TJR tunnel, prosthetic joint, and dead (Supplementary Figure 2). All participants enter the model once managing OA in the community, then some progress to a 12-month TJR tunnel, followed by living the rest of their lives with a prosthetic joint. Each health state estimates annual cost, utilities, and transition probabilities. We assumed the risk of death is the same in all health states. The model did not include revisions, infections, and TJR for different joints because we assumed these clinical characteristics were equal between cohorts. We assumed GLA:D reduced the risk of TJR because 11% of GLA:D participants in Canada report being willing to have TJR before the program and being unwilling to have TJR after the program.[36] Cycle lengths were 12 months. The model was run for 20 cycles (average life expectancy at age 65 in Canada[37]). Observed costs and outcomes were used to populate the model when possible. Cost and outcomes were discounted at 3% annually to include the time preference for money and health.[25] Model parameters

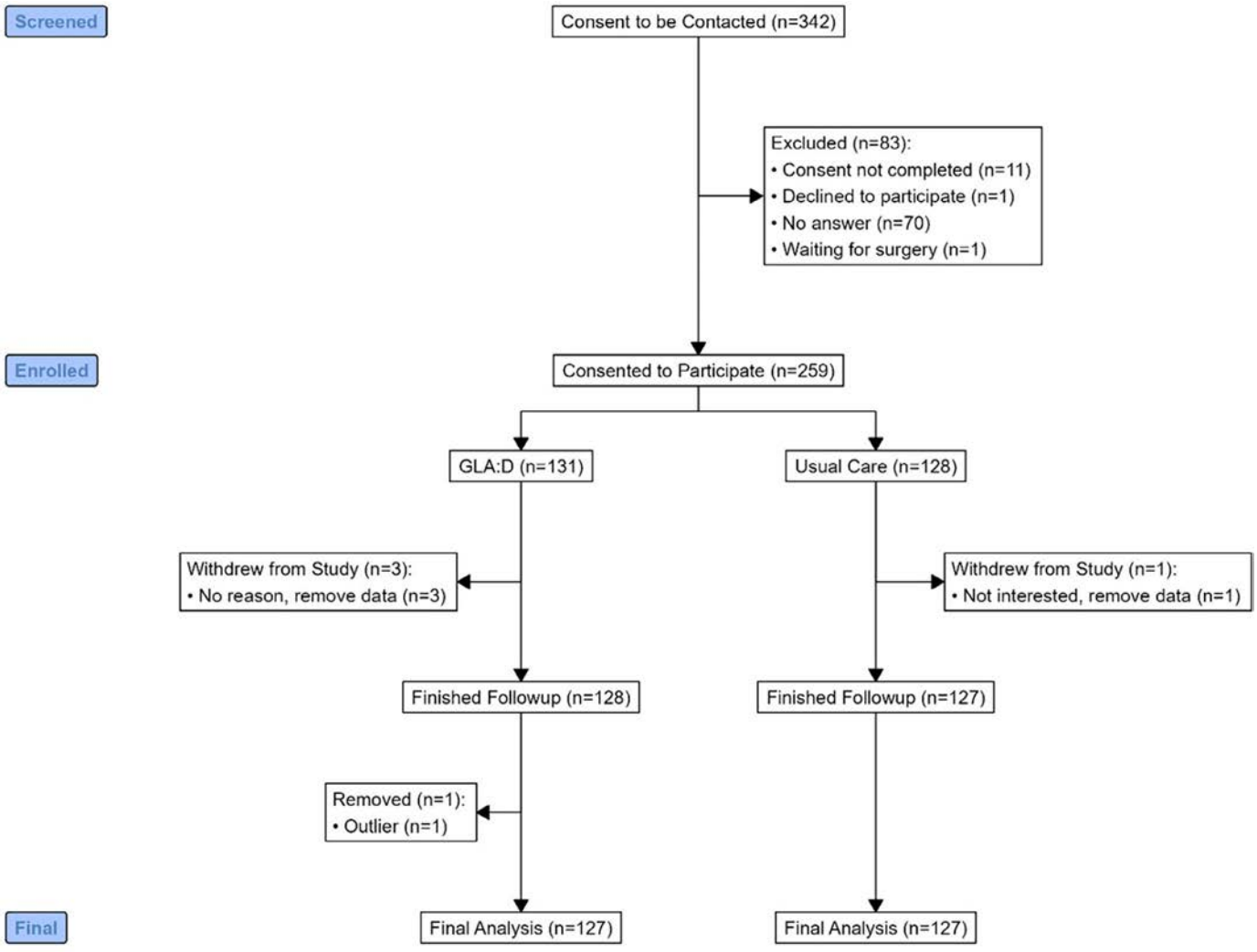


Figure 1. Participant flow diagram. GLA:D, Good Life with osteoArthritis in Denmark.

and transition probabilities are shown in Supplementary Table 1.

Probabilistic sensitivity analysis was used to characterize uncertainty in parameter estimates. A randomly selected value was chosen from the distribution of each parameter. The results stabilized after 2,000 iterations, but we conducted 10,000 iterations to align with best practices.[25]

RESULTS

Participants. A total of 254 participants were included, and 83 were excluded (Figure 1). Baseline characteristics were similar, except there were more women and worse function in the GLA:D cohort (Table 1). Most participants (96%) had private health insurance, but 29% of plans did not cover allied health care or medical devices.

Health outcomes. GLA:D and UC participants gained a mean ± SD of 0.70 ± 0.18 and 0.68 ± 0.18 QALYS over 12 months (Table 2). GLA:D participants gained 7.76 ± 19.90, 4.98 ± 21.26, 8.93 ± 19.26, and 7.18 ± 18.28 on the pain, function, QoL, and summary scores. Meanwhile, UC participants gained 0.25 ± 17.65, −2.84 ± 16.40, 1.79 ± 17.24, and −0.31 ± 14.86 on the pain, function, QoL, and summary scores. Differences in pain, function, QoL, and summary scores were statistically significant between the GLA:D and UC participants.

Costs and cost-effectiveness over 12 months. The average cost per patient to manage OA for 12 months was $1,604 and $5,035 in the GLA:D cohort and $1,683 and $3,913 in the UC cohort from an MOH and health care perspective (Table 2). GLA:D produced $6,065 (95% confidence interval [CI]: $3,648–$8,482) and $499 (95% CI: −$2,913 to $3,912) adjusted INMB in comparison to UC over 12 months from an MOH and health care perspective in the primary scenario analysis in which all cases without surgery were included (Table 3). GLA:D produced higher INMB in all scenarios from the MOH perspective, but scenarios produced conflicting results from the health care perspective. GLA:D produced positive adjusted INMB for WTP thresholds ranging from $7,000 to $30,000 (Figure 2).

Cost-effectiveness over lifetime. In the deterministic analysis, GLA:D produced an adjusted INMB of $5,957 and $618 compared to UC over a lifetime time horizon from the MOH and health care perspectives in the primary scenario analysis, in which all cases without surgery were included. In the

Table 1. Demographics of the study participants*

Variable	GLA:D, n = 127	Usual care, n = 127	*P* value[a]
Joint, n (%)			0.2
Knee	73 (57.5)	87 (68.5)	
Hip	34 (26.8)	26 (20.5)	
Knee and hip	20 (15.7)	14 (11.0)	
Female sex, n (%)	101 (79.5)	82 (64.6)	0.012
Age, mean ± SD, y	65.11 ± 7.09	63.54 ± 11.94	0.6
Body mass index, mean ± SD	30.42 ± 6.55	29.39 ± 6.42	0.2
Postsecondary education, n (%)	61 (50.4)	69 (54.8)	0.2
Married, n (%)	89 (73.6)	87 (69.0)	0.3
Symptom duration, mean ± SD, y	5.23 ± 5.79	5.82 ± 5.66	0.13
Wants a TJR n (%)	52 (44.1)	59 (47.6)	0.6
Number of comorbidities, n (%)			>0.9
0	28 (25.5)	29 (24.8)	
1	29 (26.4)	28 (23.9)	
2	25 (22.7)	30 (25.6)	
3+	28 (25.5)	30 (25.6)	
[Missing]	17	10	
Retired, n (%)	74 (61.2)	68 (54.4)	0.3
Annual household >CAD$60,000, n (%)	72 (66.1)	86 (69.4)	
Health utilities (QALY),[b] mean ± SD	0.67 ± 0.19	0.69 ± 0.21	0.13
HOOS-12/KOOS-12,[c] mean ± SD			
Pain	48.24 ± 16.81	49.58 ± 17.65	0.3
Function	54.44 ± 19.58	60.01 ± 20.92	0.024
Quality of life	36.15 ± 16.97	39.79 ± 18.81	0.072
Summary	46.29 ± 15.49	49.86 ± 17.40	0.054

* GLA:D, Good Life with osteoArthritis in Denmark; HOOS-12, 12-item Hip Injury and Osteoarthritis Outcome Score; KOOS-12, 12-item Knee Injury and Osteoarthritis Outcome Score; QALY, quality-adjusted life years; TJR, total joint replacement.
[a] Pearson's chi-squared test; Fisher's exact test; Wilcoxon rank sum test.
[b] European Quality of Life 5-Dimension five-level version is used to calculate QALYs ranging from −0.148 (worst) to 0.949 (best) quality of life.
[c] Scores in the HOOS-12 and KOOS-12 range from 0 (worst) to 100 (best).

Table 2. Unadjusted health outcomes and costs over 12 months*

Variable	GLA:D			Usual care			*P* value
	n	Mean (±SD)	Median (IQR)	n	Mean (±SD)	Median (IQR)	
Unadjusted health outcomes							
QALYs[a]	68	0.70	0.18	91	0.68	0.18	0.60
HOOS-12/KOOS-12[b]							
Pain	83	7.76	19.90	99	0.25	17.65	<0.01
Function	79	4.98	21.26	99	−2.84	16.40	<0.01
Quality of life	84	8.93	19.26	101	1.79	17.24	<0.01
Summary	76	7.18	18.28	93	−0.31	14.84	<0.01
Unadjusted costs, CAD$							
MOH costs	127	1,604 (3,789)	265 (967)	127	1,683 (3,890)	380 (877)	
Health care costs	127	5,035 (7,778)	1,985 (5,273)	127	3,913 (5,483)	1,881 (3,808)	
MOH NMB	68	19,357 (7,468)	21,383 (5,621)	91	18,841 (7,048)	21,461 (8,824)	
Health care NMB	68	15,350 (11,427)	18,858 (10,552)	91	16,275 (8,358)	19,339 (11,158)	

* GLA:D, Good Life with osteoArthritis in Denmark; HOOS-12, 12-item Hip Injury and Osteoarthritis Outcome Score; KOOS-12, 12-item Knee Injury and Osteoarthritis Outcome Score; MOH, Ministry of Health; NMB, net monetary benefit; QALY, quality-adjusted life years.
[a] European Quality of Life 5-Dimension five-level version is used to calculate QALYs ranging from −0.148 (worst) to 0.949 (best) quality of life.
[b] Scores in the HOOS-12 and KOOS-12 range from 0 (worst) to 100 (best).

probabilistic analysis, GLA:D produced an INMB of $3,990 and $1,039 compared to UC over a lifetime time horizon from the MOH and health care perspectives (Supplementary Table 2).

Supplementary Figure 3 shows the incremental difference in cost and outcomes between GLA:D and UC over a lifetime time horizon from the MOH perspective in the probabilistic analysis (10,000 iterations). The likelihood of GLA:D being cost-effective at a WTP threshold of $30,000/QALY was 51.8% and 51.3% from an MOH and health care perspective (Figure 3).

DISCUSSION

Over 12 months, using a threshold of WTP of $30,000 per QALY, GLA:D produced higher INMB than UC in the MOH perspective. The INMB was highly uncertain from the health care perspective. Similar results were observed over a lifetime time horizon but with a high degree of uncertainty. Our results suggest that health systems can generate more value for money by funding first-line OA treatments like GLA:D. Monetary benefits may decline over time or when privately funded costs are considered.

Our results can be generalized to people managing symptomatic hip and/or knee OA in the community. Our sample had similar demographics to GLA:D participants in Alberta.[36] However, our sample was slightly younger, more educated, and had higher incomes than a population-level cohort in Ontario, Canada,[38] which reflects similar demographic differences observed in the general population of Alberta and Ontario. Demographic differences are unlikely to impact the generalizability of the health effects of a structured exercise and education program. It is possible that higher incomes in Alberta may contribute to higher out-of-pocket costs and the increased uncertainty that we observed in the estimates from the health care perspective.

Modest net monetary gains were generated from the combination of small health improvements in GLA:D and small cost differences. GLA:D showed statistically significant improvements in pain, function, QoL, and summary scores compared to UC over 12 months. However, clinical outcomes did not meet the minimally important change, the smallest change that a patient perceives as clinically important (14.9 points[39]). In contrast, there were no differences in health utilities between cohorts. We

Table 3. Summary of incremental net monetary benefit over 12 months from the MOH and health care perspective*

Scenario	Sample size	INMB, CAD$	95% confidence interval, CAD$	SD, CAD$
MOH perspective				
All cases, no surgery[a]	231	6,065	3,648 to 8,482	1,223
All cases	254	3,414	341 to 6,487	1,556
Complete cases, no surgery	139	6,360	3,549 to 9,171	1,421
Complete cases	154	4,657	1,217 to 8,096	1,740
Health care Perspective				
All cases, no surgery	231	499	−2,913 to 3,912	1,731
All cases	255	−1,581	−5,906 to 2,774	2,195
Complete cases, no surgery	139	2,068	−1,913 to 6,049	2,195
Complete cases	154	1,264	−4,183 to 6,711	2,756

* MOH is the primary perspective. INMB, incremental net monetary benefit; MOH, Ministry of Health.
[a] Primary scenario analysis.

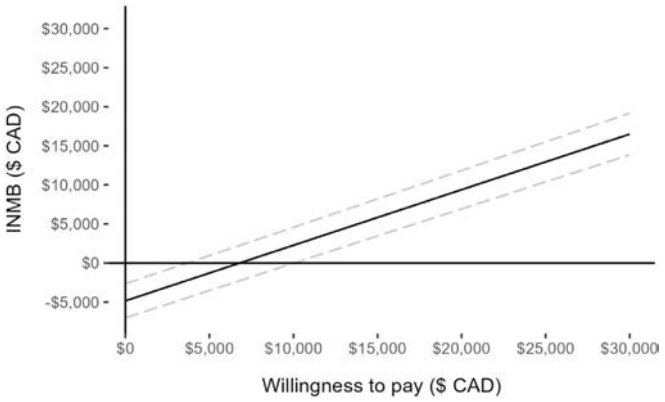


Figure 2. Incremental net monetary benefit at a range of willingness to pay thresholds (Ministry of Health perspective, complete cases). INMB, incremental net monetary benefit.

calculated the health utility difference from baseline to 12 months and still did not find statistical differences between cohorts. This suggests the small clinical effects observed with the disease-specific QoL measure and no effects observed with the generic QoL measure were related to the measure's sensitivity in our sample instead of the methods used to score each measure (slope versus area under the curve calculation). Controlling for baseline differences produced a statistically significant INMB in the MOH perspective over 12 months, whereas unadjusted QALY and cost differences were not statistically significant between cohorts. Different results were also seen when a broader range of costs was considered from the health care perspective. Participants used many health care services from the private marketplace to manage their hip and/or knee OA symptoms, producing

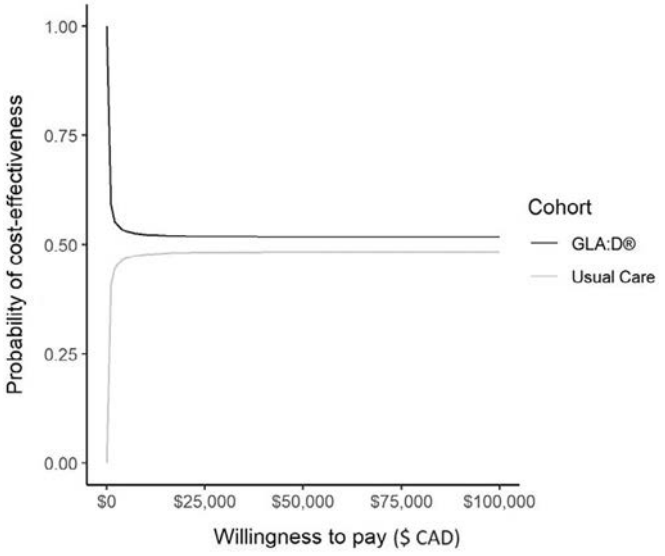


Figure 3. Cost-effectiveness acceptability curve comparing GLA:D to usual care at willingness to pay thresholds ranging from CAD$0 to CAD$100,000 (Ministry of Health, all cases, no surgery). GLA:D, Good Life with osteoArthritis in Denmark.

large out-of-pocket and private insurance costs. The average participant went to three massage, two chiropractors, and one acupuncture appointments during the study. We took a person-centered approach by including these costs from the health care perspective, although many of these services are not recommended by clinical guidelines. Guidelines focus on risks and benefits, but people living with OA may access services for a variety of reasons, including symptom modification, treatment plan adherence, and psychosocial support. Value for money was reduced by including participants' and private insurance costs, but this finding is probably trivial to decision-makers who do not bear these costs. Meanwhile, participants must find value in using adjunct therapies even when our results from the health care perspective were negligible.

Our real-world results align with previous trial-based economic evaluations that showed exercise interventions were cost-effective at conventional WTP thresholds compared to UC.[10] Four studies included in this systematic evaluated group exercise and education programs compared to UC, and all results showed the intervention was either cost-effective[40,41] or cost-saving.[42,43] We also found similar results to an economic evaluation conducted alongside an RCT by Skou et al,[11] although our intervention was shorter (12 weeks versus 8 weeks), and our sample population was from the real world. Standardized education and exercise therapy programs like GLA:D may produce even greater net monetary benefits if these programs help participants avoid surgery. A recent Markov model suggests that first-line OA treatments could be cost-saving if surgery is avoided for two to five years.[44] We cautiously estimated that GLA:D reduced surgical risk by 11%, and we found similar results when we tested our assumption by dropping the estimate to 0% in an unreported secondary analysis. Our assumption is likely underestimated because RCTs have shown that 68% of total knee replacement candidates[12] and 44% of total hip replacement candidates[45] randomized to an exercise intervention did not proceed to surgery after long-term follow-up.

Our results show important policy considerations in the Canadian health system. Canada's universal publicly insured health system provides 100% coverage for a narrow basket of services delivered in-hospital or by doctors, but almost all community-based OA services delivered by allied health professionals are funded privately. Estimates suggest that 70% of all health care costs in Canada are publicly funded,[46] but our results showed that the MOH funded only 10% of the total health care costs for managing hip and knee OA in the community. Participants paid 59% of health care costs out-of-pocket, which means access to first-line OA care in Canada is based on the ability to pay instead of need. Reduced access to first-line OA care will likely impact people with low socioeconomic status, who also have higher OA prevalence.[47] Including first-line OA treatments in Canada's basket of publicly insured services would be efficient and may also reduce inequitable access to proven treatments.

GLA:D is not currently funded at all in many settings in Alberta, and in the settings where it is covered, the cost is about $350 to $400. At this low threshold of WTP, these findings support the policy recommendation to fund GLA:D.

This is the first economic evaluation comparing a standardized education and exercise therapy program to UC in the real world. Real-world data maximize generalizability and provide the most relevant evidence to inform resource decisions. However, real-world data have several limitations. Patients seek care based on their preferences, which may bias the sample recruited in a real-world study. Regression techniques allowed us to control for baseline differences between cohorts, but unobserved confounding or effect modification is possible. Two data collection portals were necessary to reduce the burden on respondents, but they could increase the amount of missing data. We did not see different missing and nonmissing data patterns between cohorts, so we assume data collection portals were not a primary cause of missing data. Missing data is also common in the GLA:D Canada database because email notifications are autogenerated. We added phone call reminders, which reduced missing data by 54%. We collected data every three months to limit recall bias.[48] Lastly, a small sample means our results could be produced by randomness, but the probabilistic analysis found comparable results (albeit with increased uncertainty) when we resampled our observed cost and effect distributions 10,000 times.

Future research could evaluate the sensitivity of generic QoL measures in people who are not eligible for TJR. Evaluating exercise therapy with weight management in a real-world setting could potentially deliver higher value for money because most people living with hip and knee OA also have excess body weight.[6] Runhaar et al showed that a diet and exercise intervention may not prevent OA in high-risk middle-aged women,[49] but ample research shows that first-line OA treatments are effective[3] and cost-effective[10] before joint replacement. The optimal timing of first-line OA treatments is poorly understood but could be important for maximizing clinical effects and health system resources. Lastly, assessing whether implementing first-line OA treatment programs reduces surgical risk in the real world has important implications as health systems grapple with the growing burden of OA.

In conclusion, publicly funding structured education and exercise therapy programs like GLA:D would be an efficient use of health system resources over 12 months based on the positive INMB compared to UC. The INMB of GLA:D remained positive but was less certain when modeling was used to extend results to a lifetime time horizon. The net monetary benefit was positive in three out of four scenarios but had high uncertainty when all services paid by the MOH, private insurance, and out-of-pocket were considered.

ACKNOWLEDGMENTS

Thank you to our participants; we could not have done this project without you. Josh Kowal and Bryanne Kennedy also supported recruitment and data collection for this study. Antonietta Fazio played an essential role in coordinating data extraction from the GLA:D Canada database.

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 Marshall 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. Steinmetz JD, Culbreth GT, Haile LM, et al; GBD 2021 Osteoarthritis Collaborators. Global, regional, and national burden of osteoarthritis, 1990-2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5(9):e508–e522.
2. Hunter DJ, Schofield D, Callander E. The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol 2014;10(7):437–441.
3. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27(11):1578–1589.
4. Hawker GA, Stewart L, French MR, et al. Understanding the pain experience in hip and knee osteoarthritis – an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16(4):415–422.
5. King LK, Marshall DA, Faris P, et al; BEST-Knee Research Team. Use of recommended nonsurgical knee osteoarthritis management in patients prior to total knee arthroplasty: a cross-sectional study. J Rheumatol 2020;47(8):1253–1260.
6. Mazzei DR, Whittaker JL, Kania-Richmond A, et al. Do people with knee osteoarthritis use guideline-consistent treatments after an orthopedic surgeon recommends nonsurgical care? A cross-sectional survey with long-term follow-up. Osteoarthr Cartil Open 2022;4(2):100256.
7. 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.
8. MacKay C, Hawker GA, Jaglal SB. Qualitative study exploring the factors influencing physical therapy management of early knee osteoarthritis in Canada. BMJ Open 2018;8(11):e023457.
9. Eyles JP, Hunter DJ, Bennell KL, et al; Joint Effort Initiative Members. Priorities for the effective implementation of osteoarthritis management programs: an OARSI international consensus exercise. Osteoarthritis Cartilage 2019;27(9):1270–1279.
10. Mazzei DR, Ademola A, Abbott JH, et al. Are education, exercise and diet interventions a cost-effective treatment to manage hip and knee osteoarthritis? A systematic review. Osteoarthritis Cartilage 2021;29(4):456–470.
11. Skou ST, Roos EM, Laursen M, et al. Cost-effectiveness of 12 weeks of supervised treatment compared to written advice in patients with knee osteoarthritis: a secondary analysis of the 2-year outcome from a randomized trial. Osteoarthritis Cartilage 2020;28(7):907–916.
12. Skou ST, Roos E, Laursen M, et al. Cost-effectiveness of total knee replacement in addition to non-surgical treatment: a 2-year outcome

from a randomised trial in secondary care in Denmark. BMJ Open 2020;10(1):e033495.

13. Skou ST, Roos EM. Good Life with osteoArthritis in Denmark (GLA:D™): evidence-based education and supervised neuromuscular exercise delivered by certified physiotherapists nationwide. BMC Musculoskelet Disord 2017;18(1):72.
14. Grønne DT, Hansen IR, Kongsted A, et al. GLA:D® Denmark Annual Report 2021. University of Southern Denmark; 2022. https://www.glaid.dk/pdf/Annual%20report%202021%20eng_f.pdf
15. Husereau D, Drummond M, Augustovski F, et al; CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations. Value Health 2022;25(1):3–9.
16. Liu X, Seidel JE, McDonald T, et al. Rural-urban differences in non-local primary care utilization among people with osteoarthritis: the role of area-level factors. Int J Environ Res Public Health 2022;19(11):6392.
17. *List of GLA:D® Canada Clinics - Alberta*. Bone and Joint Canada; 2023. Accessed November 5, 2023. https://gladcanada.ca/wp-content/uploads/2024/09/List-of-GLAD-Clinics-AB-Sept-25-2024-1.pdf
18. Young JJ, Perruccio AV, Veillette CJH, et al. The GLA:D® Canada program for knee and hip osteoarthritis: a comprehensive profile of program participants from 2017 to 2022. PLoS One 2023;18(8):e0289645.
19. DADOS platform website. Techna Institute. Accessed November 5, 2023. https://www.dadosproject.com/
20. Harris PA, Taylor R, Minor BL, et al; REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208.
21. *EQ-5D-5L User Guide*. EuroQol Research Foundation; 2019. Accessed November 5, 2023. https://euroqol.org/publications/user-guides
22. Xie F, Pullenayegum E, Gaebel K, et al; Canadian EQ-5D-5L Valuation Study Group. A time trade-off-derived value set of the EQ-5D-5L for Canada. Med Care 2016;54(1):98–105.
23. Alberta Population Norms for EQ-5D-5L. Health Quality Council of Alberta; 2014. Accessed November 5, 2023. https://www.hqca.ca/wp-content/uploads/2018/05/2014_EQ_5D_5L_report_FINALFINAL-1.pdf
24. Roos E. The 2012 User's Guide to Knee Injury and Osteoarthritis Outcome Score. University of Southern Denmark; 2012. https://koos.nu/koos12userguide_RC.pdf
25. Lee KM, McCarron CE, Byran S, et al. CADTH Methods and Guidelines: Guidelines for the Economic Evaluation of Health Technologies: Canada. 4th ed. Canadian Agency for Drugs and Technologies in Health; 2017. Accessed November 15, 2023. https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf
26. Alberta Health Care insurance plan: schedule of medical benefits as of 01 October 2019. Government of Alberta website. 2019. Accessed November 5, 2023. https://open.alberta.ca/publications/somb-2019-10-01
27. *DAD Resource Intensity Weights and Expected Length of Stay for CMG+ 2015*. Canadian Institute for Health Information; 2015. Accessed November 5, 2023.
28. Briggs A. Economic evaluation and clinical trials: size matters. BMJ 2000;321(7273):1362–1363.
29. Leys C, Ley C, Klein O, et al. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J Exp Soc Psychol 2013;49(4):764–766.
30. Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford University Press; 2015.
31. Claxton K. NOAHE Rounds: Estimating Cost Effectiveness Thresholds for Canadian Provinces. Network of Alberta Health Economists; 2018. https://noahe.ca/events/past/noahe-rounds/health-rounds-ii/hetar-ii-s1
32. Paulden M. Towards an evidence-based approach for CADTH's cost-effectiveness threshold. NOAHE Rounds. [Internet]. Edmonton, Canada; 2022. https://noahe.ca/download/an_evidence_based_approach_to_cadth___s_cost_effectiveness_threshold.pdf
33. Hiligsmann M, Cooper C, Guillemin F, et al. A reference case for economic evaluations in osteoarthritis: an expert consensus article from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Semin Arthritis Rheum 2014;44(3):271–282.
34. Filipović-Pierucci A, Zarca K, Durand-Zaleski I. "Markov models for health economic evaluation: the R package heemod." April 25, 2017. https://arxiv.org/abs/1702.03252
35. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019;393(10182):1745–1759.
36. Zywiel MG, McGlasson R. GLA:D® Canada 2020 Annual Report. Bone and Joint Canada; 2021. Accessed November 5, 2023. https://gladcanada.ca/wp-content/uploads/2021/10/2020-GLAD-Annual-Report-Final.pdf
37. Archived - Life expectancy at birth and at age 65, by province and territory, three-year average. Statistics Canada website. 2019. Accessed November 5, 2023. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310040901
38. Hawker GA, Wright JG, Coyte PC, et al. Differences between men and women in the rate of use of hip and knee arthroplasty. N Engl J Med 2000;342(14):1016–1022.
39. Eckhard L, Munir S, Wood D, et al. Minimal important change and minimum clinically important difference values of the KOOS-12 after total knee arthroplasty. Knee 2021;29:541–546.
40. Hurley MV, Walsh NE, Mitchell HL, et al. Economic evaluation of a rehabilitation program integrating exercise, self-management, and active coping strategies for chronic knee pain. Arthritis Rheum 2007;57(7):1220–1229.
41. Fernandes L, Roos EM, Overgaard S, et al. Supervised neuromuscular exercise prior to hip and knee replacement: 12-month clinical effect and cost-utility analysis alongside a randomised controlled trial. BMC Musculoskelet Disord 2017;18(1):5.
42. Hurley MV, Walsh NE, Mitchell H, et al. Long-term outcomes and costs of an integrated rehabilitation program for chronic knee pain: a pragmatic, cluster randomized, controlled trial. Arthritis Care Res (Hoboken) 2012;64(2):238–247.
43. Jessep SA, Walsh NE, Ratcliffe J, et al. Long-term clinical benefits and costs of an integrated rehabilitation programme compared with outpatient physiotherapy for chronic knee pain. Physiotherapy 2009;95(2):94–102.
44. Vetsch T, Taeymans J, Lutz N. Optimising the current model of care for knee osteoarthritis with the implementation of guideline recommended non-surgical treatments: a model-based health economic evaluation. Swiss Med Wkly 2023;153(153):40059.
45. Svege I, Nordsletten L, Fernandes L, Risberg MA. Exercise therapy may postpone total hip replacement surgery in patients with hip osteoarthritis: a long-term follow-up of a randomised trial. Ann Rheum Dis 2015;74(1):164–169.
46. *National Expenditure Trends 1975 to 2019*. Canadian Institute for Health Information; 2019. Accessed November 5, 2023. https://www.cihi.ca/sites/default/files/document/nhex-trends-narrative-report-2019-en.pdf

47. O'Donnell S, Lagace C, Bancej C, et al; Arthritis in Canada Editorial Board. Life with Arthritis in Canada: A Personal and Public Health Challenge. Public Health Agency of Canada; 2010. https://www.canada.ca/en/public-health/services/chronic-diseases/arthritis/life-arthritis-canada-a-personal-public-health-challenge.html
48. Petrou S, Murray L, Cooper P, et al. The accuracy of self-reported healthcare resource utilization in health economic studies. Int J Technol Assess Health Care 2002;18(3):705–710.
49. Runhaar J, van Middelkoop M, Reijman M, et al. Prevention of knee osteoarthritis in overweight females: the first preventive randomized controlled trial in osteoarthritis. Am J Med 2015;128(8):888–895.e4.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1204–1213
DOI 10.1002/acr.25564

American College of Rheumatology
Empowering Rheumatology Professionals

Moderators and Mediators of Pain and Function Outcomes in a New Service Delivery Model for Management of Knee Osteoarthritis in Primary Care: Secondary Exploratory Analysis of a Randomized Controlled Trial

Abdolhay Farivar,[1] Jocelyn L. Bowden,[1] Kim L. Bennell,[2] Andrew M. Briggs,[3] Stephen J. Bunker,[4] Rana S. Hinman,[2] Thorlene Egerton,[2] Simon D. French,[5] Marie Pirotta,[4] Karen Schuck,[1] Venkatesha Venkatesha,[6] Nicholas A. Zwar,[7] and David J. Hunter[1]

Objective. Our objective was to explore moderators and mediators influenced changes in pain and function in people with knee osteoarthritis (OA) receiving a new model of primary care service delivery (Optimizing Primary Care Management of Knee Osteoarthritis [PARTNER]), at 12 months (ACTRN: 12617001595303).

Methods. This was a secondary analyses of a cluster randomized controlled trial comparing PARTNER to the usual general practitioner–delivered care (n = 217 patients: 112 PARTNER patients and 105 usual care patients) on knee pain and function. Pain was measured using a numerical rating scale (range 0–10, with a higher score indicating more severe pain), and function was measured using the function subscale of the Knee Injury and Osteoarthritis Outcome Score (range 0–100, with a higher score indicating better outcome). Baseline variables selected as potential moderators included age, sex, body mass index, pain duration, residential state, living arrangements, education, employment status, back pain, and other joint issues. Mediation variables included physical activity, fear of movement, pain catastrophizing, OA self-management, self-efficacy, sleep, fatigue, quality of life, depression, and satisfaction.

Results. For change in pain, no moderators influenced the intervention effect. However, age moderated change in function, with intervention participants <50 years demonstrating greater functional improvement than their older counterparts, compared to the control group (50–69 years: coefficient −32.88, 95% confidence interval [CI] −45.02 to −20.74; ≥70 years: coefficient −24.28, 95% CI −36.53 to −12.02). Mediation analysis revealed significant indirect effects of overall, treatment-related, and symptom-related satisfaction on mean change in pain (−0.10, −0.06, and −0.08, respectively) and function (0.09, 0.05, and 0.07, respectively).

Conclusion. Younger PARTNER participants showed greater functional improvement compared to older age groups (moderating effect). Additionally, indirect mediation effects suggest increased satisfaction across the three satisfaction domains led to reduced knee pain and enhanced function.

INTRODUCTION

Osteoarthritis (OA), a common disabling joint disease, significantly impacts individuals and societies worldwide.[1] It is the most prevalent form of arthritis, affecting millions globally and placing a substantial burden on health care systems. The prevalence of OA, particularly knee OA, is universally increasing, likely due to escalating global obesity rates, aging populations, and physical inactivity.[1–4] Knee OA accounts for approximately 60% of the worldwide burden of OA.[2,4] The core recommended approaches to OA management are lifestyle and behavioral treatments, including patient education for self-management, therapeutic

ACTRN: 12617001595303.

Supported by an Australian Government National Health and Medical Research Council (NHMRC) Partnership Grant (APP1115720). The NHMRC had no role in the design or other components of the study. Also supported by our private health insurer partner organizations (Medibank Better Health Foundation and Bupa Australia) who declare an interest in the outcome, Arthritis Australia, Medibank Private, Good2Give, Monash University, Precedence Health Care, and HealthChange Australia. The NHMRC Centre of Research Excellence for Translational Research in Musculoskeletal Pain (APP1079078) provided additional funding and in-kind support for components of the study outside the scope of the NHMRC grant. Dr Bennell's work was supported by an NHMRC Investigator Grant (1174431). Dr Hinman's work was supported by an NHMRC Investigator Grant (2025733). Dr Pirotta's work was supported by an NHMRC Career Development Fellowship. Dr Hunter's work was supported by an NHMRC Investigator Grant. We confirm the independence of the researchers from funders and that all authors had full access to all data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

[1]Abdolhay Farivar, MD, Jocelyn L. Bowden, BHMSc, BSc(Hons), PhD, Karen Schuck, BPhty, David J. Hunter, MBBS, PhD, FRACP: Royal North Shore

SIGNIFICANCE & INNOVATIONS

- This study explored factors influencing knee pain and function in people with knee osteoarthritis (OA) following the new Optimizing Primary Care Management of Knee Osteoarthritis (PARTNER) model of service delivery.
- Participant demographic variables did not exhibit any moderating effect on the PARTNER intervention's ability to change pain in people with knee OA.
- Age was a moderator of change in function following the PARTNER intervention, with younger people in the PARTNER group demonstrating greater functional improvements than older people, when compared to usual general practitioner–delivered care.
- The majority of the impacts of the PARTNER model of service delivery on pain and function were mediated through participants' overall, treatment-related, and symptom-related satisfaction levels, highlighting the importance of ensuring responsive, person-centered, and acceptable care in people with knee OA.

exercise and physical activity, and weight loss (if required).[5–7] These treatments can be supported by the judicious use of pharmacologic interventions for additional pain management, as necessary. However, the effective delivery of these interventions in real-world care has encountered challenges in implementation because of a lack of knowledge and confidence by health professionals to deliver this type of care, a lack of services in the community to support the lifestyle and behavioral changes needed, and health system structures such as financing models.[8–10]

The new Optimizing Primary Care Management of Knee Osteoarthritis (PARTNER) model of service delivery was developed to bridge this gap between the recommended evidence-based OA care and its actual delivery in Australian primary care, using a telehealth approach.[11] The PARTNER model aimed to enhance outcomes in people with OA by improving general practitioner (GP) knowledge on current evidence-based OA care, and then providing eligible participants with knee OA access to a centralized multidisciplinary "care support team" (CST). The role of the CST was to remotely support participants in their OA self-management for 12 months. The PARTNER model was tested against usual GP care in a cluster randomized controlled trial (cRCT)[11,12] to compare changes in the two primary outcomes of self-reported knee pain and knee function at 12 months. The primary results showed that although the PARTNER model showed significant improvements in knee pain and function compared to the usual GP care, the magnitude of improvement was not clinically meaningful for pain and was uncertain for function.[13]

To understand the cRCT outcomes and their consequences and implications for research and practice more comprehensively, it is important to examine what other factors may have affected the primary outcomes. Identifying potential moderators of the PARTNER primary outcomes, for example, would enable us to concentrate and allocate resources more effectively to the subset of people with knee OA most likely to benefit from the model, particularly when resources are constrained. Moderators describe the participants' baseline characteristics that interact with the treatment to influence clinical outcomes, thus revealing the populations or situations in which the treatment may be most effective, particularly when an unexpectedly weak or inconsistent relationship exists between an intervention and an outcome.[14,15] On the other hand, mediators are variables through which the intervention affects the outcome (Supplementary Figure 1). Considering potential mediators of the intervention can provide insights into why and how an intervention operates[14] and provide a basis to refine the PARTNER model to enhance effectiveness. Therefore, this study sought to identify the moderators and mediators of the effect of the PARTNER model of service delivery on changes in knee pain and function at 12 months, compared to usual GP care.

PATIENTS AND METHODS

Study design. This is a secondary analysis of data from the PARTNER study, a pragmatic, two-arm cRCT conducted within general practices in New South Wales (NSW) and Victoria, Australia (September 2018 to December 2020). Participant recruitment was finalized in December 2019 before the onset of COVID-19, and there was no impact of the pandemic on recruitment, randomization, or baseline data collection. Written informed consent was obtained from the general practices, GPs, and patient participants. The study protocol was approved by The

Hospital, Kolling Institute, and The University of Sydney, Sydney, New South Wales, Australia; [2]Kim L. Bennell, BPhysio(Hons), PhD, Rana S. Hinman, BPhysio(Hons), PhD, Thorlene Egerton, BPhty, MPhil, PhD: Centre for Health, Exercise and Sports Medicine, The University of Melbourne, Melbourne, Victoria, Australia; [3]Andrew M. Briggs, BSc(Physiotherapy)Hons, PhD, FACP: Curtin School of Allied Health, Curtin University, Perth, Western Australia, Australia; [4]Stephen J. Bunker, RN, PhD, Marie Pirotta, MBBS, MMed, PhD, FRACGP: The University of Melbourne, Melbourne, Victoria, Australia; [5]Simon D. French, BAppSc(Chiropractic), MPH, PhD: Macquarie University, Sydney, New South Wales, Australia; [6]Venkatesha Venkatesha, BSc, MSc, PhD, Northern Sydney Local Health District, Sydney, New South Wales, Australia; [7]Nicholas A. Zwar, MBBS, MHP, PhD, FRACGP: University of New South Wales, Sydney, New South Wales, Australia, and Bond University, Gold Coast, Queensland, Australia.

Drs Farivar and Bowden are co-first authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25564).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25564.

Address correspondence via email to David J. Hunter, MBBS, PhD, FRACP, at david.hunter@sydney.edu.au.

Submitted for publication June 5, 2024; accepted in revised form April 15, 2025.

University of Sydney Human Research Ethics Committee (2016/959), prospectively registered (ACTRN: 12617001595303) and published.[11] The primary results were published[13] in 2023. This article is reported in accordance with the A Guideline for Reporting Mediation Analyses statement.[16]

Study participants. A total of 38 general practices, 71 GPs, and 217 people with knee OA (112 people in the PARTNER group and 105 people in the usual care group) were enrolled in the study,[13] although the intended sample size was 44 general practices and 572 people with OA. Recruitment was stopped early due to a delayed start, slow initial recruitment, and our funder's time frame. Eligibility criteria for participant inclusion were age 45 years or older, experiencing activity-related knee pain for a duration exceeding three months, and reporting moderate to very severe pain of 4 or more on an 11-point numerical rating scale (NRS) (higher scores indicated more severe pain). Exclusion criteria were an inability to consent in English, limited mobility requiring the use of a wheelchair or scooter, a history of (or had been booked for) knee replacement surgery targeting the specified knee, or the presence of a significant health condition affecting participation.

Randomization was performed at the level of the general practice. Participating practices were assigned to either the PARTNER group or usual care group, using a 1:1 allocation ratio in random permuted blocks and stratified by geographic location (metropolitan or regional and rural) and practice size (<4 GPs and ≥4 GPs).

The PARTNER intervention and the usual care group. Full details of the PARTNER model and its development were published elsewhere.[11,12,17] The PARTNER model intervention had a dual focus on the person with OA and their GPs. Intervention group GPs were given the opportunity to update their knowledge regarding the management of OA through a suite of online professional development activities. After visiting their GP for a knee pain–specific consultation, participants in the PARTNER arm were referred to the multidisciplinary CST. The primary role of the CST team was to offer guidance on OA management and behavior change support and to empower people to effectively self-manage their knee OA. The CST consultations were delivered remotely (eg, phone, email, mail, short message service) according to the preferences of the participants. The primary focus of the PARTNER intervention was to encourage greater uptake of therapeutic exercise (leg strengthening), physical activity, and weight loss (if required). Participants had the option of undertaking additional secondary interventions according to clinical need, including online cognitive behavioral therapy courses on pain coping,[18] sleep,[19] and depression[20] if the participant met the predetermined criteria for stepping up care,[11] and the topic areas were identified as a priority action by the person. In contrast, GPs in the usual care group did not receive any specific OA training, and their patients solely received the GP's usual care.[11]

Data collection and primary outcome variables. Data were collected online via the REDCap software or through hardcopy surveys at baseline, 6 months, and 12 months. For this analysis, data were extracted from the baseline and 12-month assessments.[13] The primary cRCT outcome variables were change at 12 months in self-reported knee pain (over last week, 11-point NRS, range from 0 to 10, higher score = worse pain) and knee function (previous week, activities of daily living [ADL] subscale of the Knee Injury and Osteoarthritis Outcome Score [KOOS], range from 0 to 100, higher score = better function).[11,17]

Primary outcomes of the cRCT. At 12 months, the PARTNER model demonstrated statistically significant reductions in pain (0.8 of 10 points, 95% confidence interval [CI] 0.2–1.4) and improvements in knee function (6.5 of 100 points, 95% CI 2.3–10.7) compared to usual care.[13] Despite the reported improvements, this change did not reach our threshold for a minimal clinically important difference, which was set at ≥1.8 NRS points for pain and ≥8 of 100 KOOS for ADL function.[13]

Potential moderators and mediators. Potential moderators were identified a priori,[11] based on clinical or theoretical rationale[21] and previous evidence associated with individual outcomes of complex OA interventions that existed at the time.[8] An additional moderator, state of residence, was added post hoc to examine if the difference in COVID-19 lockdowns in each state influenced the results.[22] Eleven baseline variables were ultimately selected (Supplementary Table 1). Categorization of the baseline variables for the moderating effect analysis was undertaken on a pragmatic approach based on the existing literature and the relative frequency of responses. Body mass index (BMI) was analyzed as both a categorical and continuous variable to enable a more detailed analysis of the results. The final variables and categories were age (<50 years, 50–69 years, and ≥70 years), sex (male or female), BMI (healthy [18.5–24.9], overweight [25–29.9], or obese [≥30]), pain duration (≤10 years or >10 years), state of residence (NSW or Victoria), living arrangements (alone or with others), educational status (with or without tertiary education), employment status (part-time job or other), inability to work due to knee problems or other health issues (yes or no), coexisting back pain (yes or no), and other coexisting joint issues (yes or no) (Supplementary Table 1).

Candidate mediators were variables measured at baseline and 12 months. Similar to the moderator selection, these variables were identified a priori,[11] and based on existing evidence at the time,[8] and their clinical and theoretical potential to influence the primary outcomes in a complex intervention.[23] The included candidate mediators, their measurement tools, and parameters are detailed in Supplementary Table 1. These included

12-month changes in self-reported physical activity, fear of movement, pain catastrophizing, ability to manage their condition, arthritis self-efficacy, sleep quality, fatigue, health-related quality of life, and depression as well as satisfaction in three domains (global, with treatment, and with change in symptoms).

Statistical analysis. All statistical analyses were conducted using Stata version 17.0 (StataCorp). The data were summarized as means, SDs (symmetric normal data), medians, ranges (skewed or ordinal data), and proportions with 95% CIs. The sample size calculations for the main trial have been reported in the main results article.[13] We did not conduct power analyses specifically for the mediation analyses. However, the sample size was comparable or slightly lower than similar studies conducted moderation and mediation analyses.[24,25]

The moderation analysis was undertaken using a general linear model to test the effect of treatment on the change in knee pain and function. The model specification included the main effects of treatment allocation (PARTNER vs usual care group), moderator variables, and their interaction effect. The model effects were adjusted for baseline pain and function scores. A univariate analysis was performed on individual candidate moderators, with a multivariable analysis undertaken if more than one candidate. The coefficient presented indicates the mean difference in the outcome (knee pain or function) between two categories of a variable after interaction with the PARTNER intervention. When the moderator was demonstrated as having a significant coefficient, it was presented as being in the 95% CI.

The mediation analysis was based on the approach done by Zhao et al[26] and involved decomposition of the total effect into the indirect effect (ie, the average causal mediation effect of treatment allocation on change in pain or function score being mediated by the variable being tested) and average direct effect (ie, the effect of treatment allocation on change in pain or function that is not passed through the mediator) (Supplementary Figure 1). Complete case data were used for the mediation analysis and performed using "mediate" and "sem" packages in Stata. The test for the significance of the mediation effect was based on a sampling distribution derived from 1,000 bootstrapping replications. Models were adjusted for baseline pain and function scores. The standardized direct, indirect, and total effects are reported along with 95% CIs and associated *P* values. Data and the analytical code are available from the authors on request.

RESULTS

Sample characteristics. Two hundred and seventeen participants had baseline data available for the moderator analysis, including 112 PARTNER participants and 105 usual care participants (Figure 1 and Supplementary Table 2). For the mediation analysis, data were available for 173 to 183 participants at 12 months, depending on the outcome measured (Supplementary Table 1). Baseline participant characteristics were similar between groups, with the exception of the number of participants reporting a BMI of ≥27 (usual care 66% vs PARTNER 47%) and those reporting back pain (usual care 30% vs

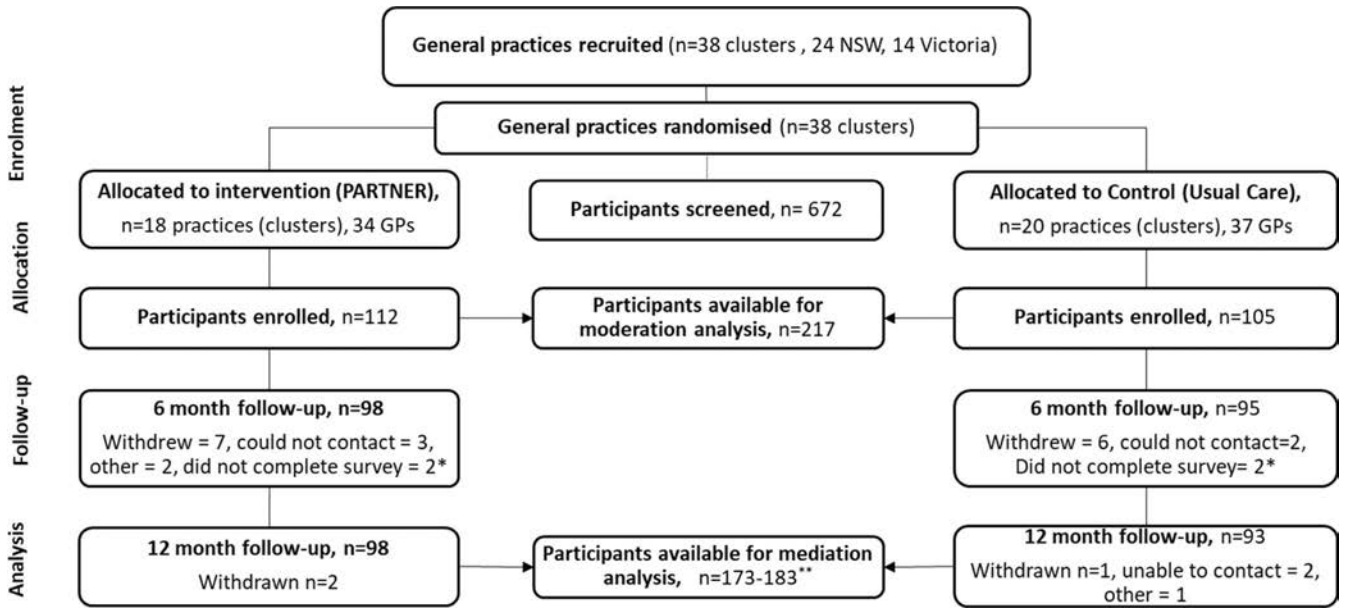


Figure 1. Consolidated Standards of Reporting Trials diagram of the PARTNER study. Participant flow through the cluster randomized controlled trial. Online screening was considered incomplete if patients did not include their contact details in the online screening tool. *Did not complete the 6-month survey but did complete the 12-month survey. **Vary depending on the outcome variable between the range. GP, general practitioner; NSW, New South Wales; PARTNER, Optimizing Primary Care Management of Knee Osteoarthritis.

PARTNER 46%). Most participants identified as female (63% PARTNER vs 57% usual care) and were aged between 50 and 69 years (both groups).

Moderation analysis. *Knee pain.* The univariate main effects and interaction effects with treatment allocation for each candidate moderator are shown in Tables 1 and 2. There was no evidence that any baseline characteristic moderated change in knee pain at 12 months (Table 1).

Knee function. Age emerged as the only moderator of change in knee function (Table 2 and Supplementary Figure 2). There was a significant interaction found between age band and knee function. It was found that the treatment effect as measured by the mean difference in KOOS ADL score between PARTNER and usual care groups for participants aged <50 years was significantly higher as compared with older age groups (Supplementary Figure 2 and Table 2). In detail, the improvement in knee function for participants aged 50 to 69 years was 32.88 points lower (95% CI 20.74–45.02) than for those under 50 years. Similarly, for participants aged 70 years and older, the improvement was 24.28 points lower (95% CI 12.02–36.53) compared to the <50 years group.

Mediation analysis. *Knee pain.* We only observed a mediating effect for variables linked to the three satisfaction measures (Tables 3 and 4). The results showed significant indirect effects of global satisfaction (–0.10, 95% CI –0.15 to –0.07), satisfaction with treatment (–0.06, 95% CI –0.10 to –0.02), and satisfaction with knee symptoms (–0.08, 95% CI –0.12 to –0.05), compared to usual care. The corresponding direct effects are shown in Table 3.

Knee function. When considering change in knee function, the three satisfaction measures showed a more pronounced influence through indirect effects on the PARTNER scores compared to the usual care group. Specifically, the effects were global satisfaction (0.09, 95% CI 0.06–0.14), satisfaction with treatment

Table 1. Univariate regression model for moderators of change in pain scores (range 0–10, higher = worse pain) from baseline to 12 months*

Characteristic	Main effect coefficient	*P* value	Interaction effect coefficient	*P* value
Age groups, y				
<50	0[a]		0[a]	
50–69	0.00 (−1.14 to 1.14)	1.000	1.95 (−0.02 to 3.92)	0.053
≥70	0.02 (−1.20 to 1.24)	0.970	1.55 (−0.43 to 3.52)	0.125
Sex				
Female	0[a]		0[a]	
Male	0.03 (−0.53 to 0.59)	0.924	−0.33 (−1.15 to 0.50)	0.440
BMI[b]	0.08 (0.02 to 0.13)	0.008	0.01 (−0.07 to 0.09)	0.833
18.5–24.9	0[a]		0[a]	
25–29.9	0.22 (−0.47 to 0.91)	0.527	0.05 (−0.94 to 1.03)	0.929
≥30	0.93 (0.19 to 1.68)	0.014	0.01 (−1.07 to 1.09)	0.987
Pain duration, y				
≤10	0[a]		0[a]	
>10	0.68 (−0.02 to 1.37)	0.057	−0.28 (−1.24 to 0.68)	0.569
State of residence				
NSW	0[a]		0[a]	
Victoria	−0.35 (−1.10 to 0.40)	0.363	0.25 (−0.73 to 1.23)	0.620
Living arrangements				
Living with others	0[a]		0[a]	
Living alone	0.76 (0.10 to 1.41)	0.024	−0.54 (−1.51 to 0.44)	0.281
Tertiary education				
No	0[a]		0[a]	
Yes	−0.66 (−1.24 to −0.09)	0.024	−0.16 (−0.99 to 0.67)	0.712
Employment as a part-time job				
No	0[a]		0[a]	
Yes	0.26 (−0.67 to 1.19)	0.586	−0.50 (−1.68 to 0.68)	0.409
Inability to work				
No	0[a]		0[a]	
Yes	1.33 (0.53 to 2.14)	0.001	−0.92 (−2.37 to 0.53)	0.215
Other joint issues				
No	0[a]		0[a]	
Yes	−0.08 (−0.68 to 0.51)	0.784	−0.06 (−1.19 to 1.07)	0.916
Back pain				
No	0[a]		0[a]	
Yes	−0.24 (−0.82 to 0.33)	0.407	0.07 (−0.74 to 0.88)	0.860

* BMI, body mass index; NSW, New South Wales.
[a] Reference group.
[b] BMI was calculated as both a continuous and a categorical variable.

Table 2. Univariate regression model for moderators of change in function scores (range 0–100, higher = better outcome) from baseline to 12 months*

Characteristic	Main effect coefficient	*P* value	Interaction effect coefficient	*P* value
Age groups, y				
<50	0[a]		0[a]	
50–69	12.30 (4.28 to 20.32)	0.003	−32.88 (−45.02 to −20.74)	<0.001
≥70	6.84 (−1.85 to 15.53)	0.123	−24.28 (−36.53 to −12.02)	<0.001
Sex				
Female	0[a]		0[a]	
Male	0.03 (−0.53 to 0.59)	0.924	−0.33 (−1.15 to 0.50)	0.440
BMI[b]	−0.76 (−1.30 to −0.21)	0.007	−0.32 (−0.98 to 0.34)	0.346
18.5–24.9	0[a]		0[a]	
25–29.9	−3.49 (−9.89 to 2.92)	0.286	4.33 (−3.93 to 12.59)	0.304
≥30	−7.96 (−15.95 to 0.04)	0.051	−4.17 (−14.08 to 5.74)	0.410
Pain duration, y				
≤10	0[a]		0[a]	
>10	−1.02 (−7.97 to 5.94)	0.774	1.87 (−7.28 to 11.01)	0.689
State of residence				
NSW	0[a]		0[a]	
Victoria	2.35 (−3.85 to 8.55)	0.458	−1.967 (−9.90 to 5.96)	0.627
Living arrangements				
Living with others	0[a]		0[a]	
Living alone	−2.98 (−8.90 to 2.93)	0.323	−3.00 (−11.29 to 5.29)	0.478
Tertiary education				
No	0[a]		0[a]	
Yes	7.01 (1.44 to 12.58)	0.014	−0.20 (−7.62 to 7.22)	0.958
Employment as a part-time job				
No	0[a]		0[a]	
Yes	5.24 (−1.57 to 12.06)	0.131	−2.00 (−10.96 to 6.95)	0.661
Inability to work				
No	0[a]		0[a]	
Yes	−20.29 (−26.02 to −14.57)	<0.001	−1.54 (−12.50 to 9.43)	0.784
Other joint issues				
No	0[a]		0[a]	
Yes	5.72 (0.34 to 11.10)	0.037	−3.18 (−11.97 to 5.61)	0.478
Back pain				
No	0[a]		0[a]	
Yes	−0.98 (−7.02 to 5.06)	0.750	−0.65 (−8.36 to 7.06)	0.868

* BMI, body mass index; NSW, New South Wales.
[a] Reference group.
[b] BMI was calculated as both a continuous and a categorical variable.

Table 3. Standardized estimates of the total, direct, and indirect effects of treatment allocation on change in pain scores at 12 months for each potential mediator considered for full mediation analysis*

Variables and measurement tools	Total effect	Direct effect	Indirect effect (95% CI)	*P* value
Physical activity levels (PASE)	−0.03	−0.04	0.00 (−0.01 to 0.01)	0.800
Fear of movement (BFMS)	−0.04	−0.05	0.01 (−0.03 to 0.04)	0.713
Pain catastrophizing (PCS)	−0.04	−0.03	−0.01 (−0.05 to 0.03)	0.592
Ability to manage their condition (EC-17)	0.03	0.03	−0.00 (−0.02 to 0.00)	0.274
Arthritis self-efficacy (ASES)	0.03	0.02	−0.01 (−0.04 to 0.01)	0.239
Sleep quality (PROMIS adult Sleep-Related Impairment SF 8a)	0.04	0.04	0.00 (−0.02 to 0.03)	0.777
Health-related quality of life (AQoL-8D)	0.03	0.02	−0.01 (−0.04 to 0.02)	0.410
Fatigue (PROMIS Fatigue SF 8a)	0.04	0.04	0.01 (−0.02 to 0.03)	0.588
Depression (PHQ-9)	0.03	0.03	−0.00 (−0.03 to 0.02)	0.873
Satisfaction, global (7-point rating scale)	−0.05	0.05	−0.10 (−0.15 to −0.01)	<0.001
Satisfaction with treatment (7-point rating scale)	−0.06	0.00	−0.06 (−0.10 to −0.02)	0.003
Satisfaction with change in symptoms (7-point rating scale)	−0.06	0.02	−0.08 (−0.12 to −0.05)	<0.001

* AQoL-8D, Assessment of Quality of Life 8 dimensions; ASES, Arthritis Self-Efficacy Scale; BFMS, Brief Fear of Movement Scale for Osteoarthritis; CI, confidence interval; EC-17, Effective Consumer Scale; PASE, Physical Activity Scale for the Elderly; PCS, Pain Catastrophizing Scale; PHQ-9, Patient Health Questionnaire 9; PROMIS, Patient-Reported Outcomes Measurement Information System; SF, Short Form.

Table 4. Standardized estimates of the total, direct, and indirect effects of treatment allocation on change in function scores at 12 months for each potential mediator considered for full mediation analysis*

Variables and measurement tools	Total effect	Direct effect	Indirect effect(95% CI)	*P* value
Physical activity levels (PASE)	0.01	0.01	−0.00 (−0.02 to 0.01)	0.736
Fear of movement (BFMS)	0.00	0.01	−0.00 (−0.05 to 0.03)	0.715
Pain catastrophizing (PCS)	0.00	−0.01	0.01 (−0.03 to 0.06)	0.619
Ability to manage their condition (EC-17)	0.00	−0.02	0.02 (−0.00 to 0.04)	0.087
Arthritis self-efficacy (ASES)	0.00	−0.03	0.03 (−0.02 to 0.07)	0.215
Sleep quality (PROMIS adult Sleep-Related Impairment SF 8a)	0.00	0.01	−0.00 (−0.04 to 0.03)	0.788
Health-related quality of life (AQoL-8D)	0.01	−0.01	0.02 (−0.03 to 0.07)	0.394
Fatigue (PROMIS Fatigue SF 8a)	0.00	0.02	−0.01 (−0.05 to 0.03)	0.579
Depression (PHQ-9)	0.00	0.00	0.00 (−0.03 to 0.04)	0.847
Satisfaction, global (7-point rating scale)	0.03	−0.06	0.09 (0.06 to 0.14)	<0.001
Satisfaction with treatment (7-point rating scale)	0.03	−0.01	0.05 (0.01 to 0.09)	0.011
Satisfaction with change in symptoms (7-point rating scale)	0.03	−0.04	0.07 (0.04 to 0.11)	<0.001

* AQoL-8D, Assessment of Quality of Life 8 dimensions; ASES, Arthritis Self-Efficacy Scale; BFMS, Brief Fear of Movement Scale for Osteoarthritis; CI, confidence interval; EC-17, Effective Consumer Scale; PASE, Physical Activity Scale for the Elderly; PCS, Pain Catastrophizing Scale; PHQ-9, Patient Health Questionnaire 9; PROMIS, Patient-Reported Outcomes Measurement Information System; SF, Short Form.

(0.05, 95% CI 0.01–0.09), and satisfaction with change in knee symptoms (0.07, 95% CI 0.04–0.11). The corresponding direct effects are shown in Table 4.

The proportion of indirect effects to direct effects and total effects are shown in Supplementary Table 3. The proportion of indirect effects to direct effects of these three satisfaction variables was always greater than one and at least 96% of the total effects of the PARTNER intervention were mediated by these satisfaction variables (Supplementary Table 3).

DISCUSSION

This study explored the potential moderators and mediators of the effects of the PARTNER model of service delivery, a telehealth service for people with knee OA, on changes in knee pain and knee function at 12 months. We found some evidence that age may moderate the effects of physical function, particularly for younger individuals, and that satisfaction may mediate changes in both knee pain and function.

In our analysis, only age emerged as a moderator of change in knee function, highlighting that adults aged less than 50 years were more likely to experience improvement in knee function with the PARTNER model than older individuals compared to the usual care group. Our results were in contrary to previous studies, in which age has not been shown to moderate OA interventions in similar populations.[25,27,28] However, our observation concerning improved responses in younger people may reflect their willingness to engage more with telehealth, either due to greater familiarity with the technology or ease of convenience of access. This is supported by an Australian cross-sectional study, which showed younger participants with hip or knee OA preferred to use mailed information packs, online education programs, and telephone helplines more than other methods due to perceived usefulness and accessibility.[29] Moreover, a UK-based study on people with psychosis showed that the younger age group had greater uptake of remotely delivered therapy, whereas older people preferred to be seen in person.[30] The use of e-consultations in primary care has also been observed to be less frequent among older people,[31] although it appears that the COVID-19 pandemic has shifted attitudes toward remotely delivered services in health, including in musculoskeletal conditions, with great satisfaction in all age groups.[32,33]

Our observation may underscore the potential utility of specifically targeting the telehealth model to younger adults, or by improving the accessibility and usability of the telehealth model to better suit older individuals. For example, this intervention was delivered only by phone. Provision of video conferencing capacity and real-time feedback of exercise in particular may improve older participants' engagement with the program, given that a recent trial in knee OA showed that video consultations with a physiotherapist for exercise, education, and physical activity were noninferior to in-person care.[34] An alternative hypothesis is that our intervention could not comprehensively address all six critical domains of intrinsic capacity (ie, locomotor capacity, psychologic capacity, cognitive capacity, hearing capacity, visual capacity, and vitality) as defined by the World Health Organization, which would be more likely to decline in older participants.[35,36] In this context, the older participants may, on average, have presented with more complex health states manifesting across intrinsic capacity domains and may be less likely to experience benefit from an intervention that targeted only OA. However, because the number of participants in the younger age group was limited (4% of all participants and 4.5% of the PARTNER group were aged <50 years), the clinical significance of our finding remains inconclusive and needs to be confirmed by further research.

Our exploration yielded no significant moderating effects of other examined parameters including sex, BMI, pain duration, state of residence, living arrangements, education or employment status, and the presence of back pain or issues in other joints. These findings are supported by two systematic reviews that found no moderating effect of age, BMI, education, pain duration, pain elsewhere, or comorbidities on changes in pain or function with exercise.[25,28] Recent registry-based data also identified that individuals with knee OA and lumbar spinal stenosis did not experience poorer outcomes of a meaningful clinical magnitude from an exercise and education intervention, compared to people with knee OA and no lumbar spinal stenosis.[37] The PARTNER model, however, is a complex multilayered approach to care, with a focus on more than just exercise, but our results are also in agreement with a secondary analysis of an RCT that evaluated a multifaceted, internet-delivered program for people with knee OA and similarly did not find any moderators at nine months.[27,38] An alternative reason for the failure to uncover more moderators could be due to having fewer than expected participants, which led to the study being underpowered. As such, the possible moderating effects of some variables may be unrecognizable.

When we examined potential mediators of the PARTNER intervention, our results showed that participant satisfaction measures (satisfaction with treatment, satisfaction with change in symptoms, and global satisfaction) demonstrated a significant mediation effect on both knee pain and knee function, and there was no other evidence of mediating effects. In other words, the pain reduction and function improvement seen with the PARTNER model mainly resulted from a higher level of satisfaction in the PARTNER participants or vice versa. The direction of causality remains unknown until proven by another clinical trial. Although participants were more satisfied with the PARTNER model than the usual care group, the overall number of participants reporting moderate to extreme satisfaction with the model (ie, score of 6 or 7 on a 7-point scale with terminal descriptors of 1 = extremely dissatisfied to 7 = extremely satisfied) was only[13] 30%. Participant satisfaction results may have been influenced by COVID-19, and the general well-being of participants during this time. Unfortunately, COVID-19 also impacted on our ability to undertake qualitative interviews to explore participant experiences with the new model, and we rely solely on our survey scores to address this question. This observation underscores the substantial influence of psychologic factors and context, particularly participant satisfaction, and emphasizes the importance of evaluation and addressing these factors in OA management strategies.

Contrary to our initial hypotheses, we did not find any mediation effect by changes in physical activity levels, fear of movement, pain catastrophizing, ability to manage their condition, arthritis self-efficacy, sleep quality, fatigue, health-related quality of life, or depression. Our findings were consistent with a secondary analysis of an RCT in which there was no association between physical activity levels and future pain or physical function during an exercise intervention for older adults with knee pain.[39] A 2023 systematic review aimed at finding potential mediators of the effects of diet and exercise on improvement of pain and function in people with OA, found some evidence of mediation effect with changes in body weight and self-efficacy.[24] In addition, a secondary analysis of an RCT identified changes in pain beliefs as a potential mediator of improvements in OA clinical outcome with exercise and manual therapy in people with hip and knee OA.[40] However, our results are not consistent with these studies. Similar to the moderators, it is feasible the mediating effects may be unrecognized because of the power of the study.

Alternatively, the results may have been influenced by low sensitivity of the measurement tools and inability to detect a significant mediation effect, or that the intervention was not powerful enough for a mediation effect. Fidelity to the intervention may also have impacted the results. Although participants in the PARTNER arm self-reported good adherence in undertaking physical activity and exercise (both 79% at 12 months) and lower adherence to weight loss (48% at 12 months),[13] we did not have any objective measures to confirm these data. The participants recruited also had a mean BMI of 28.8, which is at the lower end of what we typically report in our clinical trials in Australia (eg, >30).[34,41,42] Many of those who were overweight or obese potentially did not achieve the weight-loss gains needed to derive benefit (ie, >7% body weight).[13] Similarly, we did not specifically recruit people with poor self-efficacy or high fear of movement, and thus the intervention was not impacted by these factors.

The notable strengths of this study were its design as a secondary analysis of a robust RCT. We included a comprehensive list of potential moderators and mediators, derived from the evidence at the time, which were mostly determined a priori. However, certain limitations warrant consideration. It is plausible that some nonsignificant interactions could be attributed to statistical power constraints rather than an actual absence of effect. The main PARTNER study was underpowered, and these analyses were all exploratory; hence, prudence dictates that further research be undertaken to validate our findings. Furthermore, the multifaceted nature of the PARTNER model complicates the identification of specific treatment components interacting with each moderator or mediator variable, thus engendering a challenge in establishing causal relationships. There is also the risk of violation by unidentified confounding variables. Finally, as the recruited PARTNER participants had limited cultural and ethnic diversity, these findings cannot be applied universally.

In conclusion, our study revealed that the PARTNER model's impact on people with knee OA may be moderated by age, particularly with individuals younger than 50 years old showing greater functional improvement than older individuals. Moreover, heightened patient satisfaction, including global satisfaction, satisfaction with treatment, and satisfaction with symptoms, mediated changes in pain and function in patients with knee OA in the PARTNER group compared to participants undertaking usual GP care.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Hunter 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

Medibank Better Health Foundation, Bupa Australia, and Medibank Private had no role in the study design or in the collection, analysis, or interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Medibank Better Health Foundation, Bupa Australia, or Medibank Private.

ACKNOWLEDGMENTS

Open access publishing facilitated by The University of Sydney, as part of the Wiley - The University of Sydney agreement via the Council of Australian University Librarians.

REFERENCES

1. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet 2019; 393(10182):1745–1759.
2. Long H, Liu Q, Yin H, et al. Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the Global Burden of Disease Study 2019. Arthritis Rheumatol 2022;74(7):1172–1183.
3. Ackerman IN, Buchbinder R, March L. Global Burden of Disease Study 2019: an opportunity to understand the growing prevalence and impact of hip, knee, hand and other osteoarthritis in Australia. Intern Med J 2023;53(10):1875–1882.
4. Steinmetz JD, Culbreth GT, Haile LM, et al; GBD 2021 Osteoarthritis Collaborators. Global, regional, and national burden of osteoarthritis, 1990-2020 and projections to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Rheumatol 2023;5(9): e508–e522.
5. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage 2019;27(11):1578–1589.
6. 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 Care Res (Hoboken) 2020;72(2):149–162.
7. Gray B, Gibbs A, Bowden JL, et al. Appraisal of quality and analysis of the similarities and differences between osteoarthritis Clinical Practice Guideline recommendations: a systematic review. Osteoarthritis Cartilage 2024;32(6):654–665.
8. 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.
9. Duong V, Hunter DJ. Osteoarthritis research is failing to reach consumers. Nat Rev Rheumatol 2023;19(8):464–465.
10. Briggs AM, Houlding E, Hinman RS, et al. Health professionals and students encounter multi-level barriers to implementing high-value osteoarthritis care: a multi-national study. Osteoarthritis Cartilage 2019;27(5):788–804.
11. Hunter DJ, Hinman RS, Bowden JL, et al; PARTNER Study Team. Effectiveness of a new model of primary care management on knee pain and function in patients with knee osteoarthritis: protocol for the PARTNER study. BMC Musculoskelet Disord 2018;19(1):132.
12. Egerton T, Hinman RS, Hunter DJ, et al. PARTNER: a service delivery model to implement optimal primary care management of people with knee osteoarthritis: description of development. BMJ Open 2020; 10(10):e040423.
13. Hunter DJ, Bowden JL, Hinman RS, et al; PARTNER Study Team. Effectiveness of a new service delivery model for management of knee osteoarthritis in primary care: a cluster randomized controlled trial. Arthritis Care Res (Hoboken) 2023;75(6):1320–1332.
14. Kraemer HC, Wilson GT, Fairburn CG, et al. Mediators and moderators of treatment effects in randomized clinical trials. Arch Gen Psychiatry 2002;59(10):877–883.
15. Memon M, Cheah JH, Ramayah T, et al. Moderation analysis: issues and guidelines. JASEM 2019;3(1):1–11.
16. Lee H, Cashin AG, Lamb SE, et al; AGReMA group. A Guideline for Reporting Mediation Analyses of randomized trials and observational studies: the AGReMA statement. JAMA 2021;326(11):1045–1056.
17. Hunter DJ, Hinman RS, Bowden JL, et al; PARTNER Study Team. Correction to: effectiveness of a new model of primary care management on knee pain and function in patients with knee osteoarthritis: protocol for the PARTNER study. BMC Musculoskelet Disord 2018; 19(1):443.
18. Rini C, Porter LS, Somers TJ, et al. Automated internet-based pain coping skills training to manage osteoarthritis pain: a randomized controlled trial. Pain 2015;156(5):837–848.
19. Christensen H, Batterham PJ, Gosling JA, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight Study): a randomised controlled trial. Lancet Psychiatry 2016;3(4):333–341.
20. O'Moore KA, Newby JM, Andrews G, et al. Internet cognitive-behavioral therapy for depression in older adults with knee osteoarthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2018;70(1):61–70.
21. Pincus T, Miles C, Froud R, et al. Methodological criteria for the assessment of moderators in systematic reviews of randomised controlled trials: a consensus study. BMC Med Res Methodol 2011; 11(1):14.
22. Stobart A, Duckett S. Australia's response to COVID-19. Health Econ Policy Law 2022;17(1):95–106.
23. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health 2016;37(1):17–32.
24. Lima YL, Lee H, Klyne DM, et al. How do nonsurgical interventions improve pain and physical function in people with osteoarthritis? A scoping review of mediation analysis studies. Arthritis Care Res (Hoboken) 2023;75(3):467–481.
25. 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.
26. Zhao X, Lynch JG Jr, Chen Q. Reconsidering Baron and Kenny: myths and truths about mediation analysis. J Consum Res 2010; 37(2):197–206.
27. Lawford BJ, Hinman RS, Kasza J, et al. Moderators of effects of internet-delivered exercise and pain coping skills training for people with knee osteoarthritis: exploratory analysis of the IMPACT randomized controlled trial. J Med Internet Res 2018;20(5):e10021.

28. Quicke JG, Runhaar J, van der Windt DA, et al. Moderators of the effects of therapeutic exercise for people with knee and hip osteoarthritis: a systematic review of sub-group analyses from randomised controlled trials. Osteoarthr Cartil Open 2020;2(4):100113.
29. Ackerman IN, Bucknill A, Page RS, et al. Preferences for disease-related education and support among younger people with hip or knee osteoarthritis. Arthritis Care Res (Hoboken) 2017;69(4): 499–508.
30. Mold F, Hendy J, Lai YL, et al. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019;7(4):e13042.
31. Mehrotra A, Paone S, Martich GD, et al. Characteristics of patients who seek care via eVisits instead of office visits. Telemed J e-Health 2013;19(7):515–519.
32. Hincapié MA, Gallego JC, Gempeler A, et al. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim Care Community Health 2020;11:2150132720980612.
33. Greenfield PT, Manz WJ, DeMaio EL, et al. Telehealth can be implemented across a musculoskeletal service line without compromising patient satisfaction. HSS J 2021;17(1):36–45.
34. Hinman RS, Campbell PK, Kimp AJ, et al. Telerehabilitation consultations with a physiotherapist for chronic knee pain versus in-person consultations in Australia: the PEAK non-inferiority randomised controlled trial. Lancet 2024;403(10433):1267–1278.
35. Cesari M, Sumi Y, Han ZA, et al. Implementing care for healthy ageing. BMJ Glob Health 2022;7(2):e007778.
36. World Health Organization. Integrated care for older people (ICOPE): guidance for person-centred assessment and pathways in primary care. World Health Organization; 2019.
37. Young JJ, Kongsted A, Hartvigsen J, et al. Associations between comorbid lumbar spinal stenosis symptoms and treatment outcomes in 6,813 patients with knee or hip osteoarthritis following a patient education and exercise therapy program. Osteoarthr Cartil Open 2022;4(4):100324.
38. Bennell KL, Nelligan R, Dobson F, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. Ann Intern Med 2017;166(7):453–462.
39. Quicke JG, Foster NE, Croft PR, et al. Change in physical activity level and clinical outcomes in older adults with knee pain: a secondary analysis from a randomised controlled trial. BMC Musculoskelet Disord 2018;19(1):59.
40. Ribeiro DC, Lamb SE, Abbott JH. Mediators of the effects of exercise and manual therapy for people with knee and hip osteoarthritis: a secondary, exploratory analysis of the MOA trial. Osteoarthr Cartil Open 2024;6(1):100431.
41. Paterson KL, Bennell KL, Campbell PK, et al. The effect of flat flexible versus stable supportive shoes on knee osteoarthritis symptoms: a randomized trial. Ann Intern Med 2021;174(4):462–471.
42. Hinman RS, Campbell PK, Lawford BJ, et al. Does telephone-delivered exercise advice and support by physiotherapists improve pain and/or function in people with knee osteoarthritis? Telecare randomised controlled trial. Br J Sports Med 2020;54(13):790–797.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1214–1221
DOI 10.1002/acr.25561

Survival in Immune Checkpoint Inhibitor–Treated Patients With Rheumatoid Arthritis and Non–Small Cell Lung Cancer: An Observational Cohort Study

Deanna P. Jannat-Khah, DrPH, MSPH,[1,2] Fenglong Xie, MD, PhD,[3] Ashish Saxena, DrPH, MSPH,[2] Jeffrey R. Curtis, MD, MS, MPH,[3] and Anne R. Bass, MD[1,2]

Objective. We aimed to compare overall survival (OS) in immune checkpoint inhibitor (ICI)–treated patients with metastatic non–small cell lung cancer (mNSCLC) with pre-existing rheumatoid arthritis (RA) versus those without RA.

Methods. A retrospective cohort study using Medicare claims data was performed. Participants included patients aged ≥66 years with a diagnosis of a malignant neoplasm of lung and bronchus who initiated nivolumab, pembrolizumab, or atezolizumab between March 4, 2015 and April 11, 2019, which is after the US Food and Drug Administration (FDA) approval of ICIs for mNSCLC but before the first FDA approval for stage III disease. Survival analysis using Kaplan-Meier and adjusted Cox proportional hazard models was performed.

Results. A total of 2,732 people with mNSCLC (N = 790 RA and N = 1,942 non-RA) were in the analytic cohort. Patients with RA were more likely to be female and had more comorbidities than patients without RA. Patients with RA were more likely to be taking steroids than those without RA (63% vs 45%) but equally likely to be taking dexamethasone, usually prescribed for cancer palliation, specifically (27% vs 28%) before ICI initiation. There was no difference in OS between the RA and non-RA NSCLC Kaplan-Meier survival curves (log-rank $P = 0.08$) and in adjusted models (hazard ratio 0.92, 95% confidence interval 0.78–1.09). Male sex, having more comorbidities, and steroid dose before ICI initiation were associated with worse OS. In a sensitivity analysis omitting patients receiving baseline dexamethasone, steroid dose before ICI initiation was no longer associated with worse OS.

Conclusion. After controlling for demographics and comorbid conditions, ICI-treated patients with RA with mNSCLC had no difference in OS compared with patients without RA. After excluding patients receiving dexamethasone, steroid dose was not associated with worse OS.

INTRODUCTION

More than 130,000 Americans are diagnosed with metastatic non–small cell lung cancer (mNSCLC) each year.[1] Of these, approximately 6% (7,800) have rheumatoid arthritis (RA), even though patients with RA constitute only 0.5% of the general population.[2,3] This overrepresentation of RA among patients with mNSCLC may be due to shared risk factors, such as smoking.[4] Although 5-year survival rates in patients with mNSCLC historically have only been around 7% with traditional chemotherapy,[1] the introduction of immune checkpoint inhibitors (ICIs) has revolutionized their management and improved long-term outcomes. For example, the addition of ICI treatment to chemotherapy extends 1-year survival to 69% in patients with mNSCLC compared with 49% with chemotherapy-based regimens alone.[5] From 2015 to 2019, three ICIs were approved by the US Food and Drug Administration (FDA) for mNSCLC: pembrolizumab, nivolumab, and atezolizumab.[6] These agents work by inhibiting the negative immune regulatory molecules programmed cell death protein-1 (PD-1: pembrolizumab and nivolumab) or programmed death-ligand 1 (atezolizumab), thereby activating the immune system to combat the cancer. However, patients with RA and other autoimmune diseases were systematically excluded from ICI clinical trials.[7–9]

Supported by the Hospital for Special Surgery's Department of Medicine Discovery Grant (to Dr Jannat-Khah), and analytic infrastructure for this analysis was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant P30-AR-072583 to Dr Curtis).

[1]Hospital for Special Surgery, New York, New York; [2]Weill Cornell Medicine, New York, New York; [3]University of Alabama at Birmingham.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25561).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25561.

Address correspondence via email to Deanna P. Jannat-Khah, DrPH, MSPH, at jannatkhahd@hss.edu.

Submitted for publication September 27, 2024; accepted in revised form April 8, 2025.

SIGNIFICANCE & INNOVATIONS

- In adjusted models, patients with rheumatoid arthritis (RA) with metastatic non–small cell lung cancer (mNSCLC) had similar overall survival (OS) compared with patients without RA.
- Age, female sex, and chronic lung disease were associated with better OS whereas increased Elixhauser score and glucocorticoid dose at the time of immune checkpoint inhibitor (ICI) initiation were associated with worse OS in ICI-treated patients with mNSCLC.
- After the exclusion of dexamethasone users, steroid dose before ICI initiation was not associated with worse OS in ICI-treated patients with mNSCLC.

In the pre-ICI era, patients with NSCLC with and without autoimmune diseases treated with traditional chemotherapy had similar survival rates.[2] However, patients with autoimmune diseases might have enhanced survival when treated with ICI because of preexisting immune activation. In addition, it is notable that therapeutic agonists targeting the same molecules blocked by ICIs can be used to treat RA (eg, abatacept [cytotoxic T-lymphocyte–associated antigen] and peresolimab [PD-1]), which suggests that blocking these molecules could induce RA flares.[10,11]

Published studies provide conflicting results regarding overall survival (OS) in ICI-treated patients with cancer with RA versus without RA or another autoimmune disease. Some show better OS[12] whereas others find no difference and some suggest worse survival.[13–15] For example, a recent matched control study of 87 patients with RA versus 203 patients without RA with cancer demonstrated a nonsignificant increased risk for death (hazard ratio [HR] 1.16, 95% confidence interval [CI] 0.86–1.57).[16] However, this and all other studies on this topic have serious limitations including small sample sizes (eg, 290–349 patients),[13,16] heterogeneous cancers[16] and cancer stages,[12] and a heterogenous mix of autoimmune diseases.[12,13] Therefore, we aimed to study a large national cohort of US patients with a single autoimmune disease (RA) and a single cancer (NSCLC) and cancer stage (stage IV) to measure the association between pre-existing autoimmunity and OS in the ICI treatment setting.

PATIENTS AND METHODS

Study population. Patients included in this retrospective cohort were aged ≥66 years with two Medicare claims made by an oncologist for a diagnosis of a malignant neoplasm of lung and bronchus (*International Classification of Diseases, Ninth Revision* [*ICD-9*] 162.9 and *Tenth Revision* [*ICD-10*] C34.X) treated with an ICI. Patients had to have initiated nivolumab between March 4, 2015, and August 16, 2018; pembrolizumab between October 2, 2015, and April 11, 2019; or atezolizumab between October 18, 2015, and March 18, 2019. These dates were chosen to include the period after FDA approval for mNSCLC but before the first FDA approval for stage III disease (May 2020) or for small cell lung cancer.[17,18] These dates therefore serve as a proxy for metastatic disease and NSCLC subtype and allowed us to homogenize cancer type and stage, because there is no diagnosis code for cancer stage or subtype in Medicare claims. Patients were excluded if they were treated with durvalumab (only approved for stage III disease) or cemiplimab, approved in 2021. All patients had at least 1 year of continuous enrollment in Medicare 1 year before and 1 year after ICI initiation, unless they died. Patients with claims for other cancer types ever were excluded from the analysis (Supplementary Table 1). Patients were observed through December 31, 2019, or time of death. Line of therapy (ICI as first vs the patient having previously received chemotherapy, eg, second or more line of therapy) was determined using all available data. Sensitivity analyses were conducted using a 2-year look back to minimize misallocation as to first-line or second-line therapy.

Definition of exposure. Patients with RA were identified from the fee-for-service Medicare claims dataset from the Center for Medicare and Medicaid Services (CMS) Chronic Conditions Data Warehouse that consisted of a 100% sample of patients with RA. Patients were defined as having RA if they had two Medicare claims at least 30 days apart associated with an *ICD-9* or *ICD-10* diagnosis code for RA (714.XX, M05.*, or M06.*, excluding M06.1 and M06.4) from a rheumatologist before a diagnosis of NSCLC, plus a claim for any disease-modifying antirheumatic drug (DMARD) before initiation of an ICI.[19] DMARDs included conventional synthetic DMARDs (methotrexate, leflunomide, hydroxychloroquine, and sulfasalazine), biologic DMARDs (etanercept, adalimumab, infliximab, golimumab, certolizumab, tocilizumab, sarilumab, abatacept, and rituximab) or targeted synthetic DMARDs (tofacitinib, upadacitinib, and baricitinib).

Patients without RA included in the comparator group were drawn from the 5% random Medicare sample. Controls were excluded if they had two claims by a rheumatologist for RA or if they had any claim for any of the DMARDs listed above, for any indication, before ICI initiation.

Medicare database. The analytic dataset was composed of Medicare parts A, B, and D claims from March 2014 to December 2019 for this analysis. Demographic variables such as age, sex, and race and ethnicity were included. The Elixhauser comorbidity index was calculated with an algorithm based on *ICD-9* and *ICD-10* codes in the 12 months before ICI initiation.[20] Chronic lung disease (CLD) was defined using *ICD-9* and *ICD-10* codes for interstitial lung disease and chronic obstructive pulmonary disease (COPD), and patients with other cancer diagnoses were excluded (Supplementary Table 1). Additional

variables included prior cancer treatment characteristics, including radiation (yes/no) and chemotherapy, kinase inhibitors, opioid treatment (yes/no), and antiangiogenic therapy (defined as chemotherapy yes/no) (Supplementary Table 1). National Drug Code numbers and Healthcare Common Procedure Codes were used to identify rheumatic disease medications and cancer treatments. Oral steroid use during the 91 days before ICI initiation was calculated by determining the total cumulative dosage and dividing it by the number of days prescribed and then converted into average daily prednisone equivalents. The year of ICI initiation was the calendar year for the first claim of the ICI treatment regimen. The Elixhauser comorbidity index was created using the SAS macro for calculating the 31-item index, and a modified version was created by omitting chronic pulmonary disease, solid tumor without metastasis, and RA/collagen as these items were either included in the model or were used to create the cohort.[20–22]

Primary outcome. The primary outcome was OS. Death was determined from the recorded date of death from Medicare files. Time from ICI initiation to date of death or censoring (December 31, 2019) was used in survival analysis.

Statistical methods. Frequencies, means, SDs, medians, and interquartile ranges (IQRs) were calculated to compare patients with and without RA. Comparisons were made using chi-square tests, *t*-tests, Wilcoxon rank-sum tests, and standard mean differences as appropriate. Standardized differences between RA and non-RA were calculated.[23] Kaplan-Meier curves were created, and log-rank tests were used to compare the curves. Adjusted Cox proportional hazard models were constructed to calculate adjusted HRs (aHRs) and 95% CIs. Models included age, race, sex, Hispanic ethnicity, previous radiation treatment ever (yes/no), the modified Elixhauser comorbidity index, DMARD use in the 91 days before ICI initiation, average daily prednisone dosage in the 91 days before ICI initiation, type of ICI treatment (PD-1 or programmed death-ligand 1), first- or second-line ICI treatment (using all available prior data), ICI monotherapy or ICI with concomitant chemotherapy, and CLD. All covariates were measured either at or before ICI initiation. No covariates that occurred after ICI initiation were included in the adjusted Cox proportional hazard models. Because there were little missing data, complete case analysis was used.

Two sensitivity analyses were conducted using a 1-year look back and a 2-year look back to determine first- or second-line ICI treatment classifications for adjusted Cox proportional hazard models. Additionally, a subanalysis consisting of an adjusted Cox proportional hazard model was conducted on the RA-only cohort and included DMARD use before ICI initiation. An additional sensitivity analysis was performed to exclude patients that had a prescription for dexamethasone in the 91 days before ICI initiation. Dexamethasone is most commonly prescribed for cancer palliation.

The proportional hazard assumptions were tested using Schoenfeld residuals. All covariates were added to the model, and backward stepwise elimination was performed. Variables were removed if they were not statistically significant, and their removal did not affect the adjusted point estimates from the model. Covariates that violated the proportional hazard assumption and were statistically significant were included in the model as a strata term. The Elixhauser comorbidity index was modified so that RA, COPD, and cancer were not included in the total calculated score included in the adjusted models. Age was on a 5-year scale, and the modified Elixhauser comorbidity score was on a 5-point scale. The average daily prednisone dosage was on a 5-mg scale. An alpha of 0.05 was used. SAS version 9.4 and Stata version 18.0 were used for analysis.

The institutional review board at the Hospital for Special Surgery determined that the study was exempt. The institutional review boards at the University of Alabama Birmingham and at FASTER Medicine approved this study. Patients and/or the public were not involved in the design, conduct, reporting, or dissemination of this research. Data are available from the CMS.

RESULTS

Figure 1 and Supplementary Figure 1 display the inclusion and exclusion criteria used to construct the analytic cohort for the RA mNSCLC and the non-RA mNSCLC cohorts, respectively. In the analytic cohort, 790 patients with RA and 1,942 patients without RA with mNSCLC were included. There was no difference in age between the two cohorts and White race was the majority (86% and 88%). There was a higher percentage of patients with RA with CLD (44.8% vs 30.3%, $P < 0.001$) (Table 1). Patients with RA were also more likely to be taking steroids than patients without RA (62.7% vs 45.0%), and their median daily steroid dose in the 91 days before ICI initiation was higher than in patients without RA (median 3.30 vs 0) (Table 1). There was no statistically significant difference in dexamethasone use between the RA and non-RA groups (Table 1). Overall, patients with RA had more comorbidities (Supplementary Table 2); the median Elixhauser score was 13 (IQR 10–17) for patients with RA compared with patients without RA, whose score was 12 (IQR 9–15) (Supplementary Table 2). There was no significant difference observed between the two groups for prior radiation treatment, specific ICI drug used, or year of ICI initiation.

Patients with and without RA had a similar frequency of receiving ICI without chemotherapy (84.7% vs 82.6%, $P = 0.20$) and of receiving ICI as first-line treatment (41.4% vs 38.9%, $P = 0.22$). Somewhat more patients with RA than non-RA ultimately received only one dose of ICI (17.7% vs 13.2%, $P = 0.003$), but the mean ± SD cumulative duration of ICI was the same in the two groups: 206 ± 291 versus 209 ± 265 days. A higher percentage of patients with RA (78.2%) died compared with patients without RA (74.6%) ($P = 0.09$) over a median time

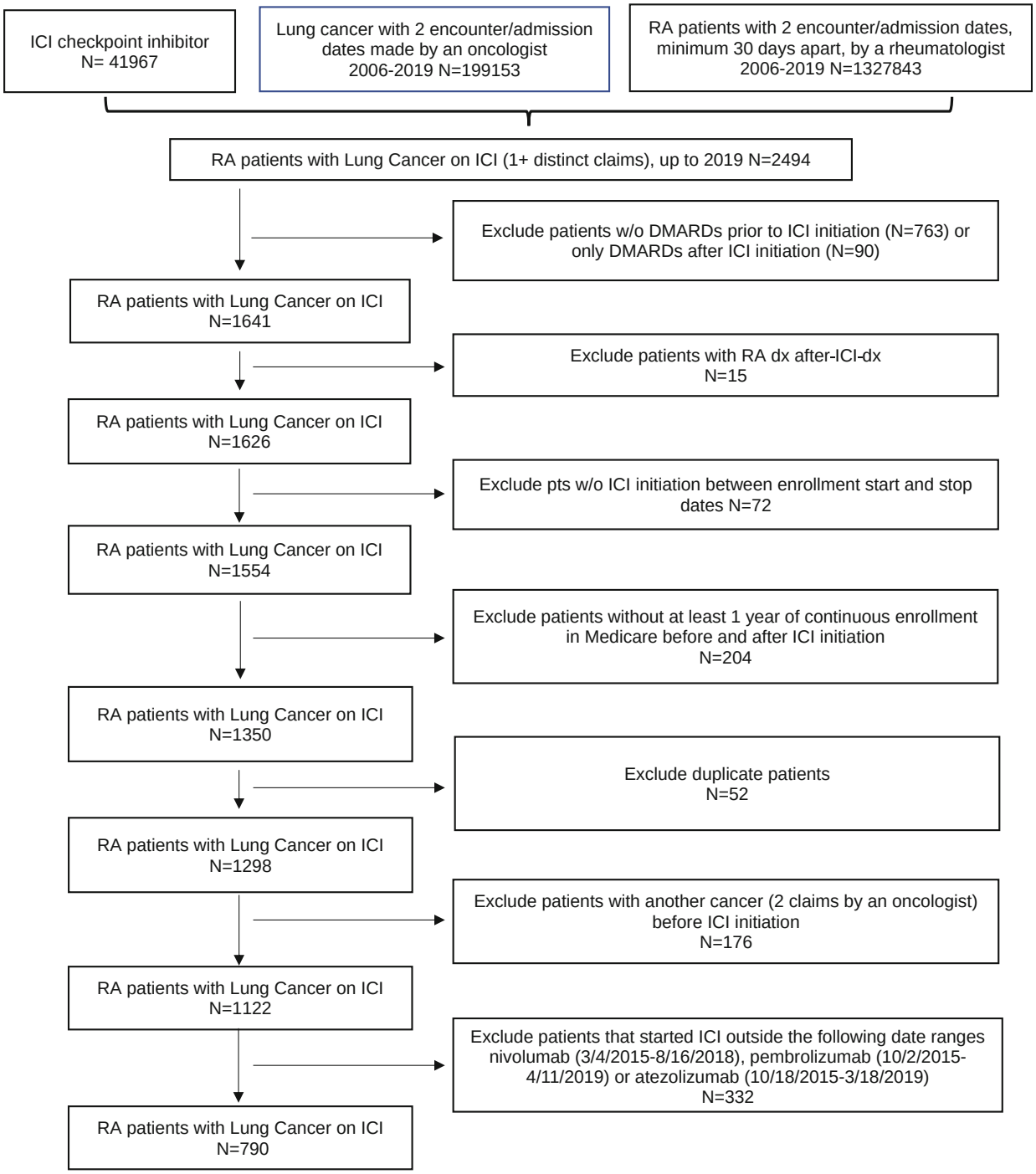


Figure 1. Exclusion cascade for the ICI-treated RA non–small cell lung cancer cohort. DMARD, disease-modifying antirheumatic drug; dx, diagnosis; ICI, immune checkpoint inhibitor; pt, patient; RA, rheumatoid arthritis.

from ICI initiation to death that was 208.5 days (IQR 79–438.5 days) for RA and 175 days (IQR 77–370 days) for non-RA (P = 0.02). The median time from ICI initiation to death or censoring was 263.5 days (IQR 94–504) for patients with RA and 286 (IQR 103–538) for patients without RA (P = 0.17).

Figure 2 displays the Kaplan-Meier curves for OS for the RA and non-RA groups from ICI initiation to death or censoring (P = 0.08). In adjusted Cox proportional hazard models for OS, there was no difference in OS between patients with and without RA (aHR 0.92, 95% CI 0.78–1.09) (Table 2). A higher Elixhauser score was associated with worse OS (aHR 1.14, 95% CI 1.06–1.22) as was glucocorticoid dose (including dexamethasone and nondexamethasone steroids) in the 91 days before ICI initiation (aHR 1.10, 95% CI 1.04–1.17)

Table 1. Patient demographics and treatment*

Factor	RA (N = 790)	Non-RA (N = 1,942)	*P* value	Standardized difference
Baseline variables				
Age, mean (SD)	74.73 (7.21)	74.62 (7.82)	0.72	−0.02
Female, n (%)	514 (65.1)	970 (49.9)	<0.001	0.31
Race, n (%)			0.003	0.19
White	693 (87.7)	1,677 (86.4)		
Black	52 (6.6)	178 (9.2)		
Asian	<11	28 (1.4)		
Other/unknown/missing	Redacted	59 (3.0)		
Hispanic	25 (3.2)	17 (0.9)	<0.001	0.20
Unknown	72 (9.1)	248 (12.8)		
Comorbidities, n (%)				
Chronic lung disease[a]	354 (44.8)	588 (30.3)	<0.001	0.30
Interstitial lung disease	135 (17.1)	100 (5.1)	<0.001	0.38683
Chronic obstructive pulmonary disease	325 (41.1)	553 (28.5)	<0.001	0.26822
Cancer treatments and medication use				
Radiation treatment, n (%)	519 (65.7)	1,328 (68.4)	0.17	0.06
Steroid use during the 91 days before ICI initiation, n (%)	495 (62.7)	873 (45.0)	<0.001	0.36
Average daily steroid dose in the 91 days before ICI initiation, median (IQR), mg prednisone equivalent	3.30 (0–10.29)	0 (0–6.59)	<0.001	−0.17
Dexamethasone use in the 91 days before ICI initiation, n (%)	209 (26.5)	552 (28.4)	0.30	0.04
Opioid use during the 91 days before ICI initiation, n (%)	486 (61.5)	1,037 (53.4)	<0.001	0.16
DMARD use during the 91 days before ICI initiation, n (%)	48 (6.1)	N/A	–	0.36
ICI treatment characteristics				
Year of ICI initiation, n (%)			0.84	0.05
2015	41 (5.2)	114 (5.9)		
2016	201 (25.4)	481 (24.8)		
2017	239 (30.3)	591 (30.4)		
2018	249 (31.5)	590 (30.4)		
2019	60 (7.6)	166 (8.5)		

* DMARD, disease-modifying antirheumatic drug; ICI, immune checkpoint inhibitor; IQR, interquartile range; N/A, not applicable; RA, rheumatoid arthritis.
[a] Includes interstitial lung disease and chronic obstructive pulmonary disease.

(Table 2). Supplementary Figure 2 displays the adjusted Kaplan-Meier curve.

As a sensitivity analysis, we excluded patients with a prescription for dexamethasone in the 91 days immediately before ICI initiation because dexamethasone is typically prescribed for cancer palliation and thus may serve as a proxy for more advanced cancer. The unadjusted Kaplan-Meier curve for OS between the RA and non-RA groups from ICI initiation to death was statistically significant, with RA having worse OS (log-rank $P = 0.01$) (Supplementary Figure 3). However, in the adjusted Cox proportional hazard model, there was no difference in OS (aHR 1.02, 95% CI 0.80–1.31) (Table 2). In addition, in this adjusted model, steroid dose was no longer associated with OS.

Two additional sensitivity analyses were conducted using a 1-year and a 2-year look back in the dataset to determine whether the patient was receiving first-line versus second-line ICI in the adjusted Cox proportional hazard models. Both produced similar results to the main analysis with no difference in OS between the two groups (Supplementary Table 3). In the subanalysis consisting of only patients with RA, a higher average daily steroid dose (aHR 1.10, 95% CI 1.03–1.18) and a higher Elixhauser score (aHR 1.16, 95% CI 1.04–1.30) were associated with worse OS (Supplementary Table 4). In Supplementary Table 5, results from a stratified model consisting of dexamethasone-only patients are reported. Average daily steroid dose (aHR 1.14, 95% CI 1.03–1.25) and a higher Elixhauser score (aHR 1.24, 95% CI 1.22–1.37) were associated with worse OS.

DISCUSSION

In this retrospective cohort study of older patients with mNSCLC enrolled in the US Medicare program, we found that patients with RA had similar OS to patients without RA after controlling for demographics, comorbidities, recent steroid exposure, and cancer treatment characteristics. Age, female sex, and CLD were found to be associated with better survival, whereas increased Elixhauser score and recent steroid exposure were associated with worse survival. However, in a sensitivity analysis in which dexamethasone-treated patients were excluded, recent steroid dose was no longer associated with OS. There are few other studies comparing OS in patients with and without RA. Van der Kooij et al found similar OS in ICI-treated patients with melanoma with an autoimmune disease versus without an autoimmune disease,[24] and a retrospective electronic health record

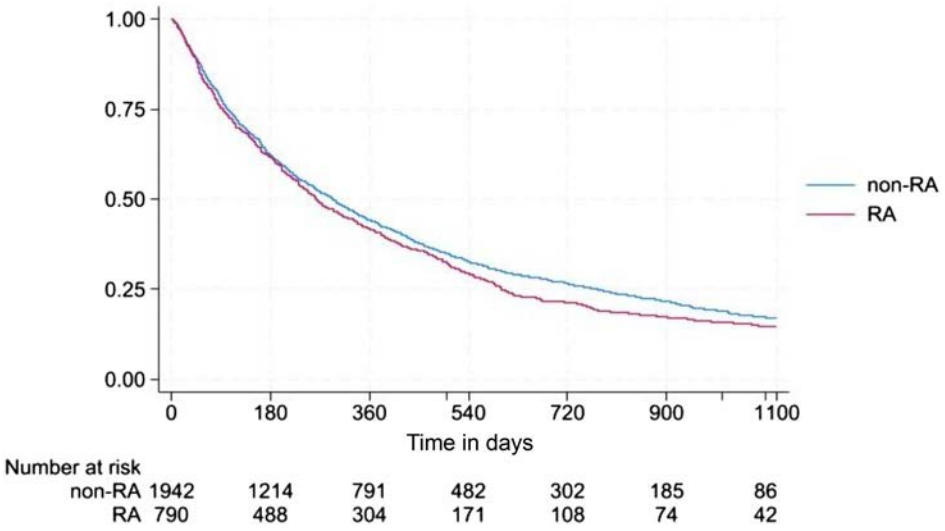


Figure 2. Kaplan-Meier curve for overall survival in immune checkpoint inhibitor–treated patients with RA non–small cell lung cancer and the non-RA comparator cohort. Log rank test *P* value = 0.08. RA, rheumatoid arthritis. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25561/abstract.

analysis by McCarter et al demonstrated a nonsignificant HR of 1.16 for death in ICI-treated patients with cancer with versus without RA. That study only included 290 individuals (87 RA and 203 non-RA), and they had multiple cancers.[16] Compared with these and other published studies, our study has a large sample size and focuses on a single cancer, single cancer stage, and a single autoimmune disease, RA. This strengthens our study's validity and avoids many of the potential biases inherent in other work.

Originally, we had hypothesized that patients with RA might be less likely to receive ICI as first-line therapy (which is associated with better OS than second-line therapy) because their oncologist might fear they would experience excessive immune-related adverse events (irAEs). However, this was not the case, as similar numbers of patients with and without RA received ICI first-line therapy (41.4% vs 38.9%, $P = 0.22$). A similar percentage of patients with and without RA were treated with dexamethasone (26.5% vs 28.4%, $P = 0.30$), a proxy for more advanced disease as this steroid is most commonly used for cancer palliation (eg, for brain metastases). This also suggests that patients with RA were not referred for ICI treatment at a later stage of their disease.

Steroid use overall (dexamethasone plus nondexamethasone) in the 3 months before ICI initiation, which was more common in the patients with RA, was found to be significantly associated with worse OS. Arbour et al also found that higher steroid use (≥10 mg of prednisone) at 30 days before ICI initiation was associated with worse OS, progression-free survival, and response rate in NSCLC.[25] However, when we performed a

Table 2. Adjusted Cox proportional hazard models for time from ICI initiation to death or follow-up*

Variable	Everyone, HR (95% CI)	Excluding baseline dexamethasone, HR (95% CI)
Non-RA	Referent	Referent
RA	0.92 (0.78–1.09)	1.02 (0.80–1.31)
Age (5-year interval)	0.94 (0.90–0.98)[a]	0.95 (0.89–1.01)
Female sex	0.84 (0.75–0.95)[a]	0.89 (0.74–1.07)
Chronic lung disease[b]	0.86 (0.76–0.98)[a]	0.80 (0.66–0.96)[a]
Average daily steroid dose in the 91 days before ICI initiation, in 5-mg prednisone equivalent increments	1.10 (1.04–1.17)[a]	1.05 (0.93–1.17)
DMARD use during the 91 days before ICI initiation	1.01 (0.69–1.47)	0.98 (0.61–1.57)
Elixhauser score (5-point interval)	1.14 (1.06–1.22)[a]	1.06 (0.96–1.17)
ICI + chemotherapy (referent to ICI as monotherapy without traditional chemotherapy)	Included as a strata term in the model given violation of the PH assumption	0.91 (0.67–1.23)

* The adjusted model includes all the covariates listed. Variables that were excluded included race, ethnicity, prior radiation treatment (yes/no), and year of ICI initiation. These were not statistically significant, and their exclusion did not affect the point estimates. CI, confidence interval; DMARD, disease-modifying antirheumatic drug; HR, hazard ratio; ICI, immune checkpoint inhibitor; PH, proportional hazard; RA, rheumatoid arthritis.
[a] Statistically significant at α = 0.05.
[b] Interstitial lung disease or chronic obstructive pulmonary disease.

sensitivity analysis in which dexamethasone users were removed, steroids were no longer associated with worse survival. Dexamethasone is typically used for cancer palliation, such as for nausea, work of breathing, or to reduce edema from brain metastases.[26,27] Thus, dexamethasone likely serves as a proxy for more advanced disease and poor survival outcomes. In the analysis of the nondexamethasone-treated patients, steroids (which were almost exclusively used in patients with RA) were no longer associated with poor survival. We included a variable for any DMARD use in the 3 months before ICI initiation in our models; however, this was not found to be statistically significant in the cohort as a whole or specifically in patients with RA. We did not analyze the association between steroid or DMARD use after ICI initiation and OS in this study, but instead focused on the variables present at baseline.

More patients with RA than without RA ultimately received only one dose of ICI treatment, suggesting that it was discontinued prematurely. One can hypothesize that patients with RA may have experienced earlier irAEs (or worsening of their RA) as a result of ICI therapy. Depending on the severity of the irAE, this may prompt either ICI pause or discontinuation,[28] which could in turn have an adverse effect on OS. IrAEs and their impact on OS as well as their prevalence among patients with autoimmune disease is an emerging area of research. Some studies find no difference in the frequency of irAEs between patients with versus without pre-existing autoimmune diseases,[29] whereas others have found an increased incidence and a faster time to irAE among patients with pre-existing autoimmune disease.[15,30,31] Ultimately, however, the patients with RA in this study did as well as those without RA from the standpoint of survival.

Perhaps surprisingly, we found that CLD (interstitial lung disease or COPD) was associated with better OS (HR 0.84, 95% CI 0.76–0.92) in adjusted models. This could be because of the widespread implementation of the Global Initiative for Chronic Obstructive Lung Disease guidelines for patients with COPD.[32] The guidelines recommend computerized tomography (CT) scans each year among former smokers with COPD.[32] Perhaps annual CT scans for those most at risk for lung cancer and close monitoring by their pulmonologists could have led to the identification of mNSCLC earlier in the disease course (although all patients had to have metastatic disease to qualify for ICI during the period studied).

Our study has several limitations. We include only patients aged ≥66 years, which could reduce generalizability. However, the median age of patients with NSCLC is 71 years, and 71.6% are aged ≥65 years.[1] As Medicare claims data do not include cancer-specific information such as cancer stage or cancer type, we limited our dataset to include people who initiated ICI during the time that they were only approved for metastatic disease to homogenize these factors. Although the Surveillance, Epidemiology, and End Results–Medicare 5% database does include information about cancer stage, the Medicare 5% sample would not have included enough patients with RA and mNSCLC to power this study.[33] Patients might have also been treated with targeted cancer therapies, specific for certain mutations, for their cancer before ICI initiation. We did not include targeted therapies in our analysis, as emerging evidence is mixed on their efficacy; however, chemotherapy and ICI has become the standard of care for NSCLC. Additionally, although we considered several timeframes to determine the line of therapy for ICI (all data and 1- and 2-year look back), we acknowledge that the line of therapy is difficult to confirm in claims data.[34] Lastly, ICI drugs may have been used "off-label" to treat patients with stage 1–3 NSCLC; however, we believe this would be a small number of patients within our cohort.

We found that ICI-treated patients with RA with mNSCLC were similar in OS compared to patients without RA after controlling for potentially confounding factors including concomitant chemotherapy and steroid use. We found similar rates of dexamethasone use, a proxy for more advanced disease, in patients with and without RA. In nondexamethasone users, steroid exposure (which was seen largely in patients with RA) was not associated with worse OS, which can provide some reassurance to rheumatologists. Although the mean duration of ICI was similar in patients with and without RA, more patients with RA received only one dose of ICI, suggesting that an early irAE or RA flare may have shortened ICI duration in a subset of these individuals. Despite this, survival was similar in the two groups. Our study suggests that in patients with RA and mNSCLC, ICI can and should be offered. However, further studies are needed to determine the safety of steroids and DMARDs after ICI initiation.

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 Jannat-Khah 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. National Cancer Institute: Surveillance, Epidemiology, and End Results Program. Cancer stat facts: lung and bronchus cancer. National Institutes of Health. Accessed October 20, 2022. https://seer.cancer.gov/statfacts/html/lungb.html
2. Khan SA, Pruitt SL, Xuan L, et al. Prevalence of autoimmune disease among patients with lung cancer: implications for immunotherapy treatment options. JAMA Oncol 2016;2(11):1507–1508.
3. Hunter TM, Boytsov NN, Zhang X, et al. Prevalence of rheumatoid arthritis in the United States adult population in healthcare claims databases, 2004-2014. Rheumatol Int 2017;37(9):1551–1557.

4. Hoovestol RA, Mikuls TR. Environmental exposures and rheumatoid arthritis risk. Curr Rheumatol Rep. 2011;13(5):431–439.
5. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al; KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. N Engl J Med 2018;378(22):2078–2092.
6. US Food & Drug Administration. Search. Accessed October 20, 2022. https://www.fda.gov/search?s=non+small+cell+lung+cancer&sort_bef_combine=rel_DESC
7. von Itzstein MS, Gonugunta AS, Mayo HG, et al. Immunotherapy use in patients with lung cancer and comorbidities. Cancer J 2020;26(6):525–536.
8. Horn L, Spigel DR, Vokes EE, et al. Nivolumab Versus Docetaxel in Previously Treated Patients With Advanced Non-Small-Cell Lung cancer: two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057). J Clin Oncol 2017;35(35):3924–3933.
9. Rzeniewicz K, Larkin J, Menzies AM, et al. Immunotherapy use outside clinical trial populations: never say never? Ann Oncol 2021;32(7):866–880.
10. Simon TA, Soule BP, Hochberg M, et al. Safety of abatacept versus placebo in rheumatoid arthritis: integrated data analysis of nine clinical trials. ACR Open Rheumatol 2019;1(4):251–257.
11. Tuttle J, Drescher E, Simón-Campos JA, et al. A phase 2 trial of peresolimab for adults with rheumatoid arthritis. N Engl J Med 2023;388(20):1853–1862.
12. Ghanem P, Murray JC, Marrone KA, et al. Improved lung cancer clinical outcomes in patients with autoimmune rheumatic diseases. RMD Open 2023;9(4):e003471.
13. Jacob S, Rahbari K, Tegtmeyer K, et al. Lung cancer survival in patients with autoimmune disease. JAMA Netw Open 2020;3(12):e2029917.
14. Khozin S, Walker MS, Jun M, et al. Real-world outcomes of patients with advanced non-small cell lung cancer (aNSCLC) and autoimmune disease (AD) receiving immune checkpoint inhibitors (ICIs). J Clin Oncol 2019;37(15_suppl):110.
15. Danlos FX, Voisin AL, Dyevre V, et al. Safety and efficacy of anti-programmed death 1 antibodies in patients with cancer and pre-existing autoimmune or inflammatory disease. Eur J Cancer 2018;91:21–29.
16. McCarter KR, Wolfgang T, Arabelovic S, et al. Mortality and immune-related adverse events after immune checkpoint inhibitor initiation for cancer among patients with pre-existing rheumatoid arthritis: a retrospective, comparative, cohort study. Lancet Rheumatol 2023;5(5):e274–e283.
17. US Food and Drug Administration. FDA approves nivolumab plus ipilimumab for first-line mNSCLC (PD-L1 tumor expression ≥1%). Published May 15, 2020. Accessed July 3, 2024. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-nivolumab-plus-ipilimumab-first-line-mnsclc-pd-l1-tumor-expression-1
18. US Food and Drug Administration. FDA approves nivolumab plus ipilimumab and chemotherapy for first-line treatment of metastatic NSCLC. FDA. Published May 27, 2020. Accessed July 3, 2024. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-nivolumab-plus-ipilimumab-and-chemotherapy-first-line-treatment-metastatic-nsclc
19. Chung CP, Rohan P, Krishnaswami S, et al. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. Vaccine 2013;31(suppl 10):K41–K61.
20. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005;43(11):1130–1139.
21. University of Manitoba. Concept: Elixhauser Comorbidity Index. Updated February 3, 2023. Accessed February 6, 2025. http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1436
22. Turner K. Elixhauser Comorbidity Index macro code ICD-10. University of Manitoba. Published May 11, 2006. Accessed February 6, 2025. http://mchp-appserv.cpe.umanitoba.ca/Upload/SAS/_ElixhauserICD10.sas.txt
23. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum 2012. Accessed May 5, 2025. https://support.sas.com/resources/papers/proceedings12/335-2012.pdf
24. van der Kooij MK, Suijkerbuijk KPM, Aarts MJB, et al. Safety and efficacy of checkpoint inhibition in patients with melanoma and preexisting autoimmune disease: a cohort study. Ann Intern Med 2021;174(5):641–648.
25. Arbour KC, Mezquita L, Long N, et al. Impact of baseline steroids on efficacy of programmed cell death-1 and programmed death-ligand 1 blockade in patients with non-small-cell lung cancer. J Clin Oncol 2018;36(28):2872–2878.
26. Haywood A, Duc J, Good P, et al. Systemic corticosteroids for the management of cancer-related breathlessness (dyspnoea) in adults. Cochrane Database Syst Rev 2019;2(2):CD012704.
27. Ryken TC, McDermott M, Robinson PD, et al. The role of steroids in the management of brain metastases: a systematic review and evidence-based clinical practice guideline. J Neurooncol 2010;96(1):103–114.
28. Hoa S, Laaouad L, Roberts J, et al. Preexisting autoimmune disease and immune-related adverse events associated with anti-PD-1 cancer immunotherapy: a national case series from the Canadian Research Group of Rheumatology in Immuno-Oncology. Cancer Immunol Immunother 2021;70(8):2197–2207.
29. Smith MH, Bass AR. Arthritis after cancer immunotherapy: symptom duration and treatment response. Arthritis Care Res (Hoboken) 2019;71(3):362–366.
30. Cortellini A, Buti S, Santini D, et al. Clinical outcomes of patients with advanced cancer and pre-existing autoimmune diseases treated with anti-programmed death-1 immunotherapy: a real-world transverse study. Oncologist 2019;24(6):e327–e337.
31. Lusa A, Alvarez C, Saxena Beem S, et al. Immune-related adverse events in patients with pre-existing autoimmune rheumatologic disease on immune checkpoint inhibitor therapy. BMC Rheumatol 2022;6(1):64.
32. Global Initiative for Chronic Obstructive Lung Disease. 2023 GOLD Report. Accessed December 21, 2023. https://goldcopd.org/2023-gold-report-2/
33. National Cancer Institute: Division of Cancer Control & Population Sciences. Brief Description of the SEER-Medicare database. National Institutes of Health. Updated May 16, 2019. Accessed July 3, 2024. https://healthcaredelivery.cancer.gov/seermedicare/overview/
34. Li B, Jin J, Guo D, et al. Immune checkpoint inhibitors combined with targeted therapy: the recent advances and future potentials. Cancers (Basel) 2023;15(10):2858.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1222–1228
DOI 10.1002/acr.25554

American College of Rheumatology
Empowering Rheumatology Professionals

Gastrointestinal Perforation as a Safety Concern Among Patients With Rheumatoid Arthritis Receiving JAK Inhibitor Therapy: A Systematic Review and Network Meta-Analysis

Thipsukhon Sathapanasiri,[1] Manuel Meraz,[1] Hamraz Mokri,[2] Sajesh K. Veettil,[3] Karn Wijarnpreecha,[4] Naomi Schlesinger,[1] Arthur Kavanaugh,[5] and Nathorn Chaiyakunapruk[6]

Objective. Gastrointestinal perforation (GIP) is a rare and life-threatening safety concern associated with JAK inhibitors (JAKi). We aimed to review the evidence regarding the risk of GIP associated with the use of JAKi in patients with rheumatoid arthritis (RA) using a systematic review and network meta-analysis approach.

Methods. A comprehensive literature search was conducted in PubMed, Embase, Cochrane Central Register of Controlled Trials, and ClinicalTrials.gov through August 2024. Included were randomized controlled trials (RCTs) comparing JAKi with other comparators in adult patients with RA (age ≥18 years) and reports of GIP. Risk ratios (RRs) with 95% confidence intervals (CIs) were estimated using a random-effects model. Surface under the cumulative ranking curves (SUCRA) were used to rank interventions.

Results. A total of 23 RCTs involving 20,023 patients were included, with a median follow-up time of 24 weeks. The overall incidence of GIP among JAKi-treated patients was 0.19% (95% CI 0.10–0.35%), with 24 events occurring out of 12,430 patients. Pairwise meta-analysis showed that the risk of GIP among patients taking JAKi was not significantly increased compared to that in those taking conventional synthetic disease-modifying antirheumatic drugs (csDMARDs) (RR 1.02; 95% CI 0.41–2.56; I^2 = 0.0%). The results of the network meta-analysis are consistent with the pairwise meta-analysis findings. Compared to csDMARDs, there was no statistically significant increase in GIP risk with JAKi (RR 0.83; 95% CI 0.37–1.84; P = 0.64) without inconsistency (P = 0.55). The SUCRA of JAKi (50.8%) and biologic DMARDs (77.0%) indicated a low risk of GIP.

Conclusion. JAKi were not associated with an increased risk of GIP compared to csDMARDs in patients with RA.

INTRODUCTION

Rheumatoid arthritis (RA) is a common systemic immune-mediated disease characterized by inflammatory arthritis, resulting in symmetric, polyarticular pain and swelling, primarily affecting the small joints of the hands and feet.[1] Treatment options for RA are diverse,[2] with JAK inhibitors (JAKi) becoming an increasingly important option. These agents inhibit the JAK/STAT pathway, which is critical in inflammatory processes. This pathway involves several key cytokines, such as interleukin (IL)–2, IL-6, IL-12, IL-15, IL-23, and interferons.[3]

Four oral JAKi are approved for RA treatment by the US Food and Drug Administration and European Medicines Agency: baricitinib,[4,5] tofacitinib,[6,7] upadacitinib,[8,9] and filgotinib[10] are associated with serious side effects, including infections, malignancies, major adverse cardiovascular events, and thrombosis.[4–10] Recent reports highlight rare but serious gastrointestinal perforation (GIP), with a nationwide cohort study showing an incidence of 2.1 per 1,000 person-years.[11] Disruption of IL-6 signaling, which is crucial for maintaining intestinal epithelial integrity and mucosal barrier function, may contribute to the risk of GIP.[12] Given the emerging evidence of GIP associated with

[1]Thipsukhon Sathapanasiri, MD, Manuel Meraz, Naomi Schlesinger, MD: University of Utah, Salt Lake City; [2]Hamraz Mokri, Msc: Erasmus University Rotterdam, Rotterdam, The Netherlands; [3]Sajesh K Veettil, PhD: IMU University and Taylor's University, Kuala Lumpur, and Taylor's University, Selangor, Malaysia; [4]Karn Wijarnpreecha, MD, MPH: University of Arizona College of Medicine and Banner University Medical Center, Phoenix; [5]Arthur Kavanaugh, MD: University of California-San Diego, La Jolla; [6]Nathorn Chaiyakunapruk, PharmD, PhD: University of Utah, Salt Lake City, and IDEAS Center, Veterans Affairs Salt Lake City Healthcare System, Salt Lake City, Utah.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25554).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25554.

Address correspondence via email to Nathorn Chaiyakunapruk, PharmD, PhD, at Nathorn.Chaiyakunapruk@utah.edu.

Submitted for publication December 5, 2024; accepted in revised form March 31, 2025.

SIGNIFICANCE & INNOVATIONS

- This study represents the first network meta-analysis to specifically focus on the risk of gastrointestinal perforation (GIP) associated with JAK inhibitors (JAKi) in patients with rheumatoid arthritis (RA).
- The study provides an evidence-based evaluation of the gastrointestinal safety of JAKi by pooling data from 23 randomized controlled trials, encompassing 20,023 patients, which enhances the generalizability of the findings.
- Our results suggest that JAKi therapy does not significantly increase the risk of GIP compared to other therapeutic options, adding valuable insights to the safety profile of JAKi and informing clinical practice guidelines for RA management.
- The study emphasizes the importance of continued pharmacovigilance and monitoring in diverse patient populations, particularly as JAKi usage grows in clinical settings.

JAKi and the lack of a clear summary of evidence on this risk, this study aims to perform a systematic review and network meta-analysis (NMA) to evaluate the risk of GIP associated with JAKi among patients with RA compared to other treatment options.

MATERIALS AND METHODS

Study design. This review was conducted following the Methodological Expectations of Cochrane Intervention Reviews. The protocol for this review has been registered with the International Prospective Register of Systematic Reviews, with the registration number CRD42024605359. The reporting of this review adhered to the guidelines set forth by the Preferred Reporting Items for Systematic Reviews and Network Meta-Analyses (PRISMA-NMA).[13]

Data source and search strategy. To identify relevant articles, a comprehensive search was conducted across multiple bibliographic databases, including PubMed, Embase, the Cochrane Central Register of Controlled Trials, and ClinicalTrials.gov, through August 2024, with no language restrictions applied. Only randomized controlled trials (RCTs) were considered for inclusion. Search strategies were developed based on the Population, Intervention, Comparator, and Study Design framework, using a combination of indexing terms (MeSH in PubMed and Emtree in Embase) and free-text keywords. The search included terms for "rheumatoid arthritis" AND "Janus kinase inhibitors" AND safety terms including "adverse event or safety" or "gastrointestinal perforation" AND terms for RCTs. Full search strategies are shown in Supplementary Table S1.

Study selection, data extraction, and quality assessment. Only RCTs were included in this review if they met the following inclusion criteria: (1) participants were adults (aged ≥18 years) diagnosed with RA; (2) the study compared one of the four oral JAKi including tofacitinib, baricitinib, upadacitinib, or filgotinib, with any comparator other than these four medications; and (3) the study reported on the outcome of GIP. Two reviewers (TS and MM) independently extracted data from each eligible study using a standardized data extraction form. The following data were collected: author; title; year of publication; study location; conflict of interest statements; baseline characteristics of the population (such as age and sex); types of interventions and comparators; details regarding uncertainty analysis; follow-up duration; baseline use of conventional synthetic disease-modifying antirheumatic drugs (csDMARDs), glucocorticoids, and nonsteroidal anti-inflammatory drugs (NSAIDs); history of inflammatory bowel disease; smoking and alcohol consumption; and instances of GIP reported in the studies. Additional relevant characteristics may be included as they arise from the data. Two independent reviewers (TS and HM) assessed the risk of bias using the Revised Cochrane Risk of Bias Tool for randomized trials (RoB 2.0).[14] A third reviewer (NC) adjudicated any disagreements between the two reviewers.

Outcome of interest, data synthesis, and statistical analysis. The primary outcome of interest is the occurrence of GIP events following treatment with one of the four JAKi including tofacitinib, baricitinib, upadacitinib, or filgotinib in patients with RA. A pairwise meta-analysis was conducted to pool risk ratios (RRs) across studies using a random-effects model.[15] In our NMA, each treatment will be represented as a distinct node within the network. Heterogeneity was assessed for both the direct meta-analysis and within the network using the Cochran Q test and I^2 statistics ($P < 0.10$ or $I^2 \geq 50\%$).[16] To evaluate the consistency of the network, a global consistency test will be conducted to ensure agreement between direct and indirect evidence.[17] Additionally, transitivity was evaluated by examining the distribution of clinical and methodologic variables that could influence the outcomes of interest. The ranking of interventions will be determined using the surface under the cumulative ranking (SUCRA) values.[18] Comparison-adjusted funnel plots will be generated to assess potential publication bias for comparisons.[19] We reported RRs as effect estimates alongside the corresponding 95% confidence intervals (CIs).

Prespecified subgroup analyses were conducted based on several clinical factors, including age groups (<50 years vs ≥50 years), types of JAKi and the mechanism of JAKi (pan-JAKi and selective JAK-1 inhibitors). Additionally, sensitivity analyses excluded studies with a high risk of bias and those with small sample sizes (below the 25th percentile).[20] All analyses will be performed using Stata version 18 (StataCorp), with a P value of <0.05 being considered statistically significant.

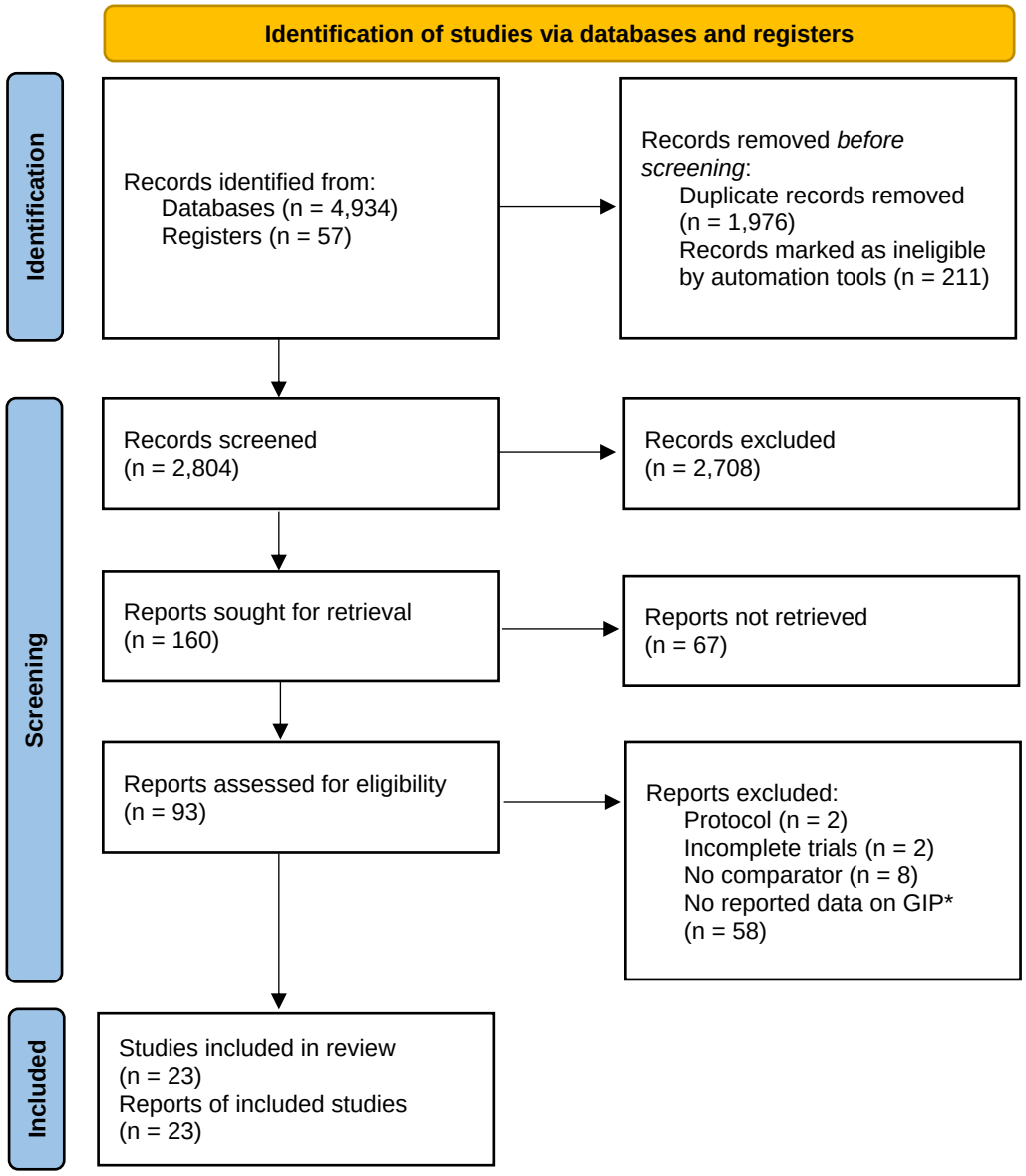
The certainty of evidence from NMA was assessed using the Confidence in Network Meta-Analysis (CINeMA) online software (https://cinema.ispm.unibe.ch/#rob [accessed November

2024]). The certainty of evidence was classified into four levels: very low, low, moderate, and high. Each outcome was graded based on six domains: within-study bias, reporting bias, indirectness, imprecision, heterogeneity, and incoherence.[21,22] Two independent reviewers (TS and SV) performed the assessment of certainty of evidence, and for cases in which discrepancies could not be resolved through discussion, a third reviewer (NC) was consulted to reach a final decision.

RESULTS

Study selection. Following a literature search using the specified search strategies (Supplementary Table S1), a total of 4,991 records were identified from various databases and registries. After conducting screening and eligibility assessments, we included 23 studies, encompassing more than 20,023 patients, in our NMA. These studies reported the incidence of GIP, enabling comprehensive comparisons between treatments. The search results and the PRISMA flowchart are presented in Figure 1.

Study characteristics. The NMA comprised a total of 23 studies with 20,023 patients, including four JAKi: tofacitinib[23–28] (6 studies, 5,748 patients), baricitinib[29–32] (4 studies, 1,079 patients), upadacitinib[33–42] (10 studies, 3,154 patients), and filgotinib[43–45] (3 studies, 2,089 patients) compared to four comparators including placebo, methotrexate, tumor



GIP; gastrointesnal perfora on.

Figure 1. Preferred Reporting Items for Systematic Reviews flow diagram. GIP, gastrointestinal perforation.

necrosis factor inhibitors (TNFi) adalimumab and etanercept, and abatacept. Most studies employed two-arm comparisons, whereas four studies used a three-arm design.[25,29,36,43] Placebo was the most common comparator, followed by methotrexate, TNFi, and, in one study, abatacept.[38]

Most studies permitted the use of concomitant csDMARDs, including methotrexate, sulfasalazine, leflunomide, hydroxychloroquine, and chloroquine, as well as glucocorticoids at a dosage of ≤10 mg/day of prednisolone equivalent, along with concomitant NSAIDs use. In studies that allowed the use of csDMARDs and had placebo as a comparator, we analyzed csDMARDs as a comparator instead of placebo, resulting in csDMARDs being the most common comparator overall.

The populations studied in these trials had a mean age ranging from 48.8 to 61.4 years and were diagnosed with active RA. The median follow-up time across all studies was 24 weeks, with a range from 12 weeks to 5.5 years. A summary of all comparisons is presented in Supplementary Table S2, whereas key characteristics of these trials are detailed in Table 1. Additional information regarding all eligible studies, including the number of patients, study populations, and interventions, can be found in Supplementary Tables S3 and S4.

Risk of bias. Using the RoB2.0[14], 60%, 7%, and 33% of the studies were classified as having low risk, some concerns, and high risk of bias, respectively (see Supplementary Table S5). Among the five domains evaluated, baseline imbalance and missing outcome data were the two most common reasons for potential bias (see Supplementary Figure S1).

Effects of JAKi on GIP. Pairwise meta-analyses comparing JAKi and csDMARDs were conducted by excluding studies without events (see Supplementary Figure S2). A total of 23 studies involving 20,023 participants and 29 reported events were analyzed. The incidence of GIP was as follows: JAKi, 1.06 per 1,000 person-years; biologic DMARDs (bDMARDs), 0.45 per 1,000 person-years; csDMARDs, 0.32 per 1,000 person-years; and placebo, 0 events. Additionally, the incidence of GIP for JAKi monotherapy was 1.46 per 1,000 person-years, and for JAKi with background csDMARDs, the incidence of GIP was 1.00 per 1,000 person-years.

A network map illustrating GIP is provided in Figure 2. A global inconsistency test revealed no evidence of inconsistency in treatment effects for the outcome ($P = 0.55$). Additionally, transitivity was assessed by comparing the distributions of age, concomitant csDMARD use, glucocorticoid and NSAIDs use, history of smoking and alcohol consumption, and history of diverticulitis or GIP. These assessments indicated no evidence of intransitivity (see Supplementary Table S4). Overall, there were no statistically significant differences in the risk of GIP between JAKi and

Table 1. Characteristics of the 23 included studies*

Study group, author, published year	Treatment	No. of patients receiving JAKi/total patients	Study duration	concomitant csDMARDs
ORAL Surveillance, Ytterberg SR, et al[23] 2022	TOF vs TNFi	2,911/4,362	5.5 y	Yes
ORAL Scan, van der Heijde D, et al[24] 2019	TOF vs PBO	637/797	24 mo	Yes
ORAL Standard, van Vollenhoven RF, et al[25] 2012	TOF vs PBO, ADA	405/717	12 mo	Yes
ORAL Strategy, Fleischmann R, et al[26] 2017	TOF vs ADA	760/1,146	12 mo	Discontinued MTX in TOF monotherapy arm
Tanaka Y, et al[27] 2015	TOF vs PBO	265/317	12 wk	No
Lee EB, et al[28] 2014	TOF vs MTX	770/956	24 mo	Discontinued MTX in TOF monotherapy arm
SELECT-BEYOND, Genovese MC, et al[33] 2018	UPA vs PBO	451/620	12 wk	Yes
SELECT-EARLY, van Vollenhoven R, et al[34] 2020	UPA vs MTX	631/945	48 wk	No
SELECT-MONOTHERAPY, Smolen JS, et al[35] 2019	UPA vs MTX	432/649	14 wk	Yes
SELECT-NEXT, Burmester GR, et al[36] 2018	UPA vs PBO	440/661	12 wk	Yes
SELECT-COMPARE, Fleischmann R, et al[37] 2019	UPA vs PBO, ADA	651/1,629	26 wk	Yes
SELECT-CHOICE, Rubbert-Roth A, et al[38] 2020	UPA vs ABA	303/612	24 wk	Yes
SELECT-SUNRISE, Kameda H, et al[39] 2020	UPA vs PBO	148/197	12 wk	Yes
Zeng X, et al[40] 2021	UPA vs PBO	169/338	12 wk	Yes
Fleischmann R, et al[41] 2022	UPA vs PBO	40/59	12 wk	Yes
BALANCE II, Genovese MC, et al[42] 2016	UPA vs PBO	249/299	12 wk	Yes
RA-BEAM, Taylor PC, et al[29] 2017	BAR vs PBO, ADA	487/1,305	52 wk	Yes
RA-BEACON, Genovese MC, et al[30] 2016	BAR vs PBO	351/527	24 wk	Yes
Tanaka Y, et al[31] 2016	BAR vs PBO	96/145	12 wk	Yes
Li Z, et al[32] 2020	BAR vs PBO	145/290	12 wk	Yes
Combe B, et al[43] 2020	FIL vs PBO, ADA	955/1,755	52 wk	Yes
FINCH II, Genovese MC, et al[44] 2019	FIL vs PBO	301/449	24 wk	Yes
FINCH III, Westhovens R, et al[45] 2020	FIL vs MTX	833/1,249	52 wk	Yes

* ABA, abatacept; ADA, adalimumab; BAR, baricitinib; csDMARD, conventional synthetic disease-modifying antirheumatic drug; FIL, jilgotinib; JAKi, JAK inhibitor; MTX, methotrexate; PBO, placebo; TNFi, tumor necrosis factor inhibitor; TOF, tofacitinib; UPA, Upadacitinib.

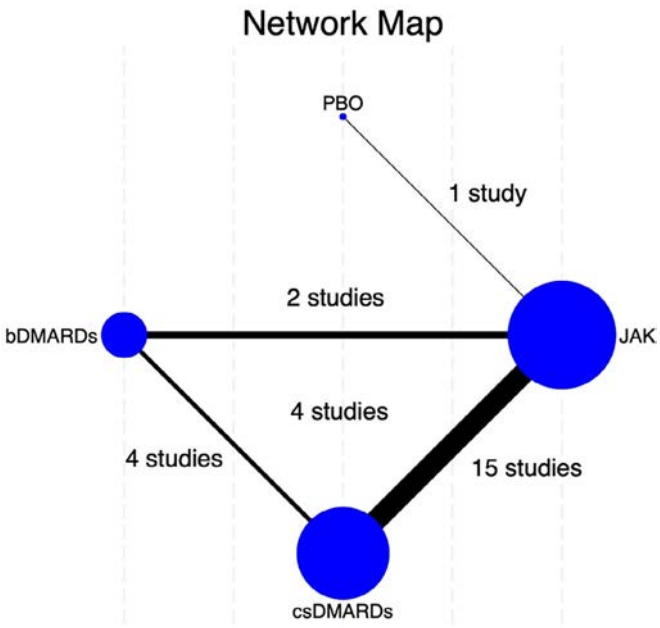


Figure 2. Network of eligible comparisons for gastrointestinal perforation outcome. bDMARD, biologic disease-modifying antirheumatic drug; csDMARD, conventional synthetic disease-modifying antirheumatic drug; PBO, placebo. Color figure can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25554/abstract.

bDMARDs compared to csDMARDs (RR 0.83, 95% CI 0.37–1.84, $P = 0.64$; see Table 2). Comparisons among all treatment interventions regarding the outcomes are detailed in Supplementary Table S6. The results of the SUCRA rankings are presented in Supplementary Figure S3 and Supplementary Table S7.

Subgroup analyses. Subgroup analyses were conducted based on age groups (<50 years and ≥50 years), types of JAKi, and mechanisms of JAKi. The effect estimates from these subgroup analyses were generally consistent with the results of the main analyses (see Supplementary Tables S8–S14).

Sensitivity analyses and publication bias. We also performed sensitivity analyses by excluding studies with a high risk of bias and those with a sample size below the 25th percentile. The effect estimates remained robust across these sensitivity analyses (see Supplementary Tables S15 and S16). Comparison-adjusted funnel plots indicated no evidence of asymmetry, suggesting no significant publication bias (see Supplementary Figure S4).

Table 2. Risk ratio of gastrointestinal perforation of interventions compared to csDMARDs*

Interventions	Risk ratio	95% CI	*P* value
JAKi	0.83	0.37–1.84	0.64
Placebo	1.38	0.05–36.97	0.85
bDMARDs[a]	0.58	0.19–1.81	0.35

* bDMARD, biologic disease-modifying antirheumatic drug; csDMARD, conventional synthetic disease-modifying antirheumatic drug; CI, confidence interval; JAKi, JAK inhibitor.
[a] bDMARDs include etanercept, adalimumab, and abatacept; all bDMARDs were administered with baseline csDMARDs.

Certainty of evidence. Applying CINeMA to the NMA, the quality of evidence for all comparisons was rated as low. Both direct and indirect evidence were also graded as low quality due to wide CIs in the imprecision section. More details on the quality of evidence can be found in Supplementary Figures S5 and S6 and Supplementary Table S17.

DISCUSSION

Our study aimed to evaluate the association between JAKi and the risk of GIP in patients with RA. The results indicate that JAKi including tofacitinib, baricitinib, filgotinib, and upadacitinib did not significantly increase the risk of GIP among patients with RA when compared to other treatment options. This finding aligns with previous studies that have suggested a relatively low incidence of GIP with JAKi,[11] despite concerns raised about their gastrointestinal safety due to their mechanism of action, which is related to the inhibition of the JAK/STAT pathway.[3] This pathway regulates several inflammatory cytokines, including IL-6, which plays a crucial role in maintaining the integrity of the intestinal mucosa.[4] The disruption of this signaling could impair mucosal repair, potentially contributing to the increased risk of GIP in certain patients. Furthermore, RA itself is associated with gastrointestinal manifestations, compounded by using concomitant medications, such as glucocorticoids and NSAIDs, which further elevate the risk of GIP.[46,47]

Despite the growing concern about GIP in patients treated with JAKi, our study found no significant increase in the risk of GIP compared to other treatment options. This is consistent with findings from Hoisnard et al[11] and Liu-Yan et al,[48] which suggest that, although there is a theoretical risk due to the mechanism of action of JAKi, it does not translate into a significantly higher incidence of GIP in real-world clinical settings or RCTs.

One of the key strengths of our study is that this is the first NMA to focus specifically on the risk of GIP associated with JAKi in patients with RA. The inclusion of a large number of trials enhances the generalizability of our findings, providing a more robust evidence base for assessing the safety of JAKi in this context. Given the growing use of JAKi in clinical practice, our findings could be valuable for the development of evidence-informed clinical practice guidelines aimed at optimizing treatment strategies for patients with RA.

However, several limitations should be considered. Firstly, the relatively small number of GIP events limits our ability to draw definitive conclusions regarding treatment-related differences in GIP risk. Given the low number of events, our findings should be interpreted with caution. Although no significant differences were observed between treatment groups, further studies with larger

sample sizes and longer follow-up periods are needed to confirm these findings and provide a more comprehensive assessment of gastrointestinal safety.

Another limitation is the relatively short follow-up period in most of the included studies, which may limit our ability to assess long-term gastrointestinal safety. Furthermore, our analysis was limited to patients with RA, meaning that the findings may not be directly applicable to other inflammatory diseases treated with JAKi. Additionally, most of the included studies did not specify how GIP was diagnosed, such as through computed tomography imaging, exploratory laparoscopy, colonoscopy, or other diagnostic methods. This lack of clarity further complicates the assessment of GIP risk.

Moreover, most of the studies did not use JAKi monotherapy, as the treatments were combined with either csDMARDs, prednisolone, or NSAIDs. This limits our ability to assess the specific impact of JAKi monotherapy on GIP risk. Both steroids and NSAIDs are known to increase the risk of gastrointestinal complications,[46,47] including perforation, and should be considered as potential confounders.

Additionally, our analysis did not account for the combined effect of NSAIDs and prednisolone use on GIP risk, as most patients continued their background therapy with NSAIDs and steroids. Given the known gastrointestinal risks of NSAIDs and steroids,[46,47] it is an important factor to consider.

Despite these limitations, our study provides valuable insights into the gastrointestinal safety of JAKi in patients with RA. Future studies should focus on conducting long-term investigations to assess the gastrointestinal safety profile of JAKi treatments over extended periods, particularly in real-world settings, to provide more comprehensive data on their long-term effects and improve clinical decision-making.

In conclusion, our results support the evidence suggesting that JAKi are not associated with GIP. However, continued pharmacovigilance is recommended to monitor GIP risk in diverse patient populations.

ACKNOWLEDGMENTS

We acknowledge the use of OpenAI's language model (available at https://www.openai.com) for assisting in improving grammar. The model was not involved in the research design, data collection, or analysis.

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 Chaiyakunapruk 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. Gravallese EM, Firestein GS. Rheumatoid arthritis - common origins, divergent mechanisms. N Engl J Med 2023;388(6):529–542.
2. Fraenkel L, Bathon JM, England BR, et al. 2021 American college of rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis Rheumatol 2021;73(7):1108–1123.
3. Nash P, Kerschbaumer A, Dörner T, et al. Points to consider for the treatment of immune-mediated inflammatory diseases with Janus kinase inhibitors: a consensus statement. Ann Rheum Dis 2021;80(1):71–87.
4. Highlights of prescribing information for olumiant (baricitinib) tablets, for oral use. US Food and Drug Administration; 2022. Accessed September 16, 2024. https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/207924s006lbl.pdf
5. Summary of product characteristics for Olumiant (baricitinib). European Medicines Agency; 2024. Accessed September 16, 2024. https://www.ema.europa.eu/en/documents/product-information/olumiant-epar-product-information_en.pdf
6. Highlights of prescribing information for XELJANZ (tofacitinib) tablets, for oral use. US Food and Drug Administration; 2018. Accessed September 16, 2024. https://www.accessdata.fda.gov/drugsatfda_docs/label/2018/203214s018lbl.pdf
7. Summary of product characteristics for Xeljanz (tofacitinib). European Medicines Agency; 2024. Accessed September 16, 2024. https://www.ema.europa.eu/en/documents/product-information/xeljanz-epar-product-information_en.pdf
8. Highlights of prescribing information for RINVOQ (upadacitinib) extended-release tablets, for oral use. US Food and Drug Administration; 2023. Accessed September 16, 2024. https://www.accessdata.fda.gov/drugsatfda_docs/label/2023/211675s015lbl.pdf
9. Summary of product characteristics for Rinvoq (upadacitinib). European Medicines Agency; 2024. Accessed September 16, 2024. https://www.ema.europa.eu/en/documents/product-information/rinvoq-epar-product-information_en.pdf
10. Summary of product characteristics for Jyseleca (filgotinib). European Medicines Agency; 2024; Accessed September 16, 2024. https://www.ema.europa.eu/en/documents/product-information/jyseleca-epar-product-information_en.pdf
11. Hoisnard L, Meyer A, Dray-Spira R, et al. Risk of gastrointestinal perforation in patients with rheumatic diseases exposed to Janus kinase inhibitors versus adalimumab: a nationwide cohort study. Arthritis Rheumatol 2024;76(9):1364–1376.
12. Alhendi A, Naser SA. The dual role of interleukin-6 in Crohn's disease pathophysiology. Front Immunol 2023;14:1295230.
13. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015;162(11):777–784.
14. Higgins JPT, Sterne JAC, Savovic J, et al. A Revised Tool for Assessing Risk of Bias in Randomized Trials. Cochrane Methods. 2016:29–31.
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177–188.
16. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557–560.
17. Dias S, Welton NJ, Caldwell DM, et al. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010;29(7–8):932–944.
18. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64(2):163–171.
19. Chaimani A, Higgins JP, Mavridis D, et al. Graphical tools for network meta-analysis in STATA. PLoS One 2013;8(10):e76654.

20. Dechartres A, Altman DG, Trinquart L, et al. Association between analytic strategy and estimates of treatment outcomes in meta-analyses. JAMA 2014;312(6):623–630.
21. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. CINeMA: an approach for assessing confidence in the results of a network meta-analysis. PLoS Med 2020;17(4):e1003082.
22. Papakonstantinou T, Nikolakopoulou A, Higgins JPT, et al. CINeMA: software for semiautomated assessment of the confidence in the results of network meta-analysis. Campbell Syst Rev 2020;16(1): e1080.
23. Ytterberg SR, Bhatt DL, Mikuls TR, et al; ORAL Surveillance Investigators. Cardiovascular and cancer risk with tofacitinib in rheumatoid arthritis. N Engl J Med 2022;386(4):316–326.
24. van der Heijde D, Strand V, Tanaka Y, et al; ORAL Scan Investigators. Tofacitinib in combination with methotrexate in patients with rheumatoid arthritis: clinical efficacy, radiographic, and safety outcomes from a twenty-four-month, phase III study. Arthritis Rheumatol 2019;71(6): 878–891.
25. van Vollenhoven RF, Fleischmann R, Cohen S, et al; ORAL Standard Investigators. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. N Engl J Med 2012;367(6):508–519.
26. Fleischmann R, Mysler E, Hall S, et al; ORAL Strategy investigators. Efficacy and safety of tofacitinib monotherapy, tofacitinib with methotrexate, and adalimumab with methotrexate in patients with rheumatoid arthritis (ORAL Strategy): a phase 3b/4, double-blind, head-to-head, randomised controlled trial. Lancet 2017;390(10093): 457–468.
27. Tanaka Y, Takeuchi T, Yamanaka H, et al. Efficacy and safety of tofacitinib as monotherapy in Japanese patients with active rheumatoid arthritis: a 12-week, randomized, phase 2 study. Mod Rheumatol 2015;25(4):514–521.
28. Lee EB, Fleischmann R, Hall S, et al; ORAL Start Investigators. Tofacitinib versus methotrexate in rheumatoid arthritis. N Engl J Med 2014; 370(25):2377–2386.
29. Taylor PC, Keystone EC, van der Heijde D, et al. Baricitinib versus placebo or adalimumab in rheumatoid arthritis. N Engl J Med 2017; 376(7):652–662.
30. Genovese MC, Kremer J, Zamani O, et al. Baricitinib in patients with refractory rheumatoid arthritis. N Engl J Med 2016;374(13):1243–1252.
31. Tanaka Y, Emoto K, Cai Z, et al. Efficacy and safety of baricitinib in Japanese patients with active rheumatoid arthritis receiving background methotrexate therapy: a 12-week, double-blind, randomized placebo-controlled study. J Rheumatol 2016;43(3):504–511.
32. Li Z, Hu J, Bao C, et al. Baricitinib in patients with rheumatoid arthritis with inadequate response to methotrexate: results from a phase 3 study. Clin Exp Rheumatol 2020;38(4):732–741.
33. Genovese MC, Fleischmann R, Combe B, et al. Safety and efficacy of upadacitinib in patients with active rheumatoid arthritis refractory to biologic disease-modifying anti-rheumatic drugs (SELECT-BEYOND): a double-blind, randomised controlled phase 3 trial. Lancet 2018; 391(10139):2513–2524.
34. van Vollenhoven R, Takeuchi T, Pangan AL, et al. Efficacy and safety of upadacitinib monotherapy in methotrexate-naive patients with moderately-to-severely active rheumatoid arthritis (SELECT-EARLY): a multicenter, multi-country, randomized, double-blind, active comparator-controlled trial. Arthritis Rheumatol 2020;72(10):1607–1620.
35. Smolen JS, Pangan AL, Emery P, et al. Upadacitinib as monotherapy in patients with active rheumatoid arthritis and inadequate response to methotrexate (SELECT-MONOTHERAPY): a randomised, placebo-controlled, double-blind phase 3 study. Lancet 2019; 393(10188):2303–2311.
36. Burmester GR, Kremer JM, Van den Bosch F, et al. Safety and efficacy of upadacitinib in patients with rheumatoid arthritis and inadequate response to conventional synthetic disease-modifying antirheumatic drugs (SELECT-NEXT): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet 2018;391(10139): 2503–2512.
37. Fleischmann R, Pangan AL, Song IH, et al. Upadacitinib versus placebo or adalimumab in patients with rheumatoid arthritis and an inadequate response to methotrexate: results of a phase III, double-blind, randomized controlled trial. Arthritis Rheumatol 2019;71(11): 1788–1800.
38. Rubbert-Roth A, Enejosa J, Pangan AL, et al. Trial of upadacitinib or abatacept in rheumatoid arthritis. N Engl J Med 2020;383(16): 1511–1521.
39. Kameda H, Takeuchi T, Yamaoka K, et al. Efficacy and safety of upadacitinib in Japanese patients with rheumatoid arthritis (SELECT-SUNRISE): a placebo-controlled phase IIb/III study. Rheumatology (Oxford) 2020;59(11):3303–3313.
40. Zeng X, Zhao D, Radominski SC, et al. Upadacitinib in patients from China, Brazil, and South Korea with rheumatoid arthritis and an inadequate response to conventional therapy. Int J Rheum Dis 2021; 24(12):1530–1539.
41. Fleischmann R, Friedman A, Drescher E, et al. Safety and efficacy of elsubrutinib or upadacitinib alone or in combination (ABBV-599) in patients with rheumatoid arthritis and inadequate response or intolerance to biological therapies: a multicentre, double-blind, randomised, controlled, phase 2 trial. Lancet Rheumatol 2022;4(6):e395–e406.
42. Genovese MC, Smolen JS, Weinblatt ME, et al. Efficacy and safety of ABT-494, a selective JAK-1 inhibitor, in a phase IIb study in patients with rheumatoid arthritis and an inadequate response to methotrexate. Arthritis Rheumatol 2016;68(12):2857–2866.
43. Combe B, Kivitz A, Tanaka Y, et al. Filgotinib versus placebo or adalimumab in patients with rheumatoid arthritis and inadequate response to methotrexate: a phase III randomised clinical trial. Ann Rheum Dis 2021;80(7):848–858.
44. Genovese MC, Kalunian K, Gottenberg JE, et al. Effect of filgotinib vs placebo on clinical response in patients with moderate to severe rheumatoid arthritis refractory to disease-modifying antirheumatic drug therapy: the FINCH 2 randomized clinical trial. JAMA 2019;322(4): 315–325.
45. Westhovens R, Rigby WFC, van der Heijde D, et al. Filgotinib in combination with methotrexate or as monotherapy versus methotrexate monotherapy in patients with active rheumatoid arthritis and limited or no prior exposure to methotrexate: the phase 3, randomised controlled FINCH 3 trial. Ann Rheum Dis 2021;80(6):727–738.
46. Harigai M. Growing evidence of the safety of JAK inhibitors in patients with rheumatoid arthritis. Rheumatology (Oxford) 2019;58(suppl 1): i34–i42.
47. Xie F, Yun H, Bernatsky S, et al. Brief report: risk for gastrointestinal perforation among rheumatoid arthritis patients receiving tofacitinib, tocilizumab, or other biologics. Arthritis Rheumatol 2016;68(11): 2612–2617.
48. Liu-Yan N, Kun Z, Cheng X, et al. Impact of upadacitinib on the risk of digestive events in patients with rheumatoid arthritis: a systematic review and meta-analysis of randomized controlled trials. Heliyon 2023;9(6):e17002.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1229–1239
DOI 10.1002/acr.25565

American College of Rheumatology
Empowering Rheumatology Professionals

Core Set of Responsive and Discriminatory Measures for Use in Pragmatic Trials of Youth With Axial Juvenile Spondyloarthritis

Timothy G. Brandon,[1] Rui Xiao,[2] Daniel J. Lovell,[3] Edward Oberle,[4] Matthew L. Stoll,[5] Nancy A. Chauvin,[6] Michael L. Francavilla,[7] Walter P. Maksymowych,[8] and Pamela F. Weiss[2]

Objective. The objective of this study was to determine a core set of measures for youth with juvenile spondyloarthritis and axial disease (axJSpA), using the juvenile arthritis working group Outcome Measures in Rheumatology framework.

Methods. This was a prospective multicenter study of youth with axJSpA. Participants (aged 8–18 years) all initiated tumor necrosis factor inhibitor (TNFi) therapy and completed questionnaires, examinations, and magnetic resonance imaging (MRI) at baseline and 12 weeks. Responsiveness and discrimination were assessed using standardized response mean (SRM) and standardized mean difference (SMD). For highly correlated ($r > |0.80|$) items within domains, larger SRM and SMD were prioritized, and minimal clinically important improvement was determined for each.

Results. Of the evaluable cohort (N = 57), 68.4% were male, and the median age was 15.3 years; 70.2% of youth treated with TNFi had clinical response (change ≥2 in patient global assessment). Although 58% had continued MRI inflammation, 77% of those patients reported moderate clinical improvement. The final axJSpA core set contained the following: Patient-Reported Outcomes Measurement Information System (PROMIS) pain interference (SRM 0.77, SMD 0.5), the sacroiliac joint inflammation score (SRM 1.02, SMD 0.52), PROMIS mobility (SRM 0.83, SMD 0.75), and patient global well-being (SRM 0.88, SMD not applicable). All overall and composite disease activity measures tested, except the physician global assessment, had high SRM and SMD. Subgroup analysis demonstrated differences by biologic sex and overweight status. Improvement in the MRI inflammation score was greater in male patients. Improvement in the PROMIS pain interference and mobility measures was greater in those with normal body mass index.

Conclusion. A set of measures was developed for youth with axJSpA.

INTRODUCTION

Approximately 15% to 25% of youth with arthritis have juvenile spondyloarthritis (JSpA) and approximately 20% of youth with JSpA have axial disease (axJSpA).[1] Even though sacroiliitis is a feature of JSpA that does not respond to standard first-line juvenile idiopathic arthritis (JIA) therapy, namely disease modifying antirheumatic drugs such as methotrexate, there is only one published clinical trial focused on youth with axial arthritis, which was ultimately a negative study perhaps due in part to the use of SpA outcomes validated in adults but not youth.[2] Despite the negative trial results, most pediatric rheumatologists firmly believe tumor necrosis factor inhibitor (TNFi) therapies are effective for axial arthritis, as evidenced by (1) an unpublished case-based survey of voting members of the Childhood Arthritis Rheumatology and Research Association (N = 369) in April 2015, which

Supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, NIH (grant 1R01-AR-074098 to Dr Weiss).

[1]Timothy G. Brandon, MPH: Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [2]Rui Xiao, PhD, Pamela F. Weiss, MD, MSCE: Perelman School of Medicine, University of Pennsylvania, and Children's Hospital of Philadelphia; [3]Daniel J. Lovell, MD, MPH: Cincinnati Children's Hospital Medical Center and University of Cincinnati, Cincinnati, Ohio; [4]Edward J. Oberle, MD: Nationwide Children's Hospital and The Ohio State University, Columbus; [5]Matthew L. Stoll, MD, PhD, MSCS: University of Alabama at Birmingham; [6]Nancy A. Chauvin, MD: The Imaging Institute, The Cleveland Clinic, Cleveland, Ohio; [7]Michael L. Francavilla, MD: Whiddon College of Medicine, University of South Alabama, Mobile; [8]Walter P. Maksymowych, MD, FRCP: University of Alberta and CARE Arthritis, Edmonton, Alberta, 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.25565).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25565.

Address correspondence via email to Pamela F. Weiss, MD, MSCE, at weisspa@chop.edu.

Submitted for publication January 7, 2025; accepted in revised form April 18, 2025.

SIGNIFICANCE & INNOVATIONS

- With many available assessment tools, including patient-reported outcomes, physician-based assessment, and imaging tools, there are no data comparing performance of outcome measures in youth with axial disease and juvenile spondyloarthritis (axJSpA) to facilitate identification of which to measure in pragmatic trials or routine clinical care.
- The most responsive and discriminative measures for youth with axJSpA in their Outcome Measures in Rheumatology domains were Patient-Reported Outcomes Measurement Information System (PROMIS) pain interference, the magnetic resonance imaging sacroiliac joint inflammation score, PROMIS mobility, and the patient global well-being visual analog scale.
- There were important subgroup differences in responsiveness, most notably in youth who were overweight versus normal weight. These standardized response mean differences were large for pain, mobility, and well-being.
- Selection of measures should take anticipated baseline disease activity, body mass index, and disease manifestations (enthesitis, peripheral arthritis) of the target population into account.

demonstrated that the use of TNFi therapy was >10 times higher for clinical scenarios of youth with peripheral arthritis and sacroiliitis compared to those with peripheral arthritis only (66% vs 6%; $P < 0.001$), and (2) inclusion of TNFi therapy as first-line therapy (after nonsteroidal anti-inflammatory drugs) in the American College of Rheumatology (ACR)/Arthritis Foundation treatment guidelines for sacroiliitis in youth with juvenile arthritis.[2]

Current JIA trials are conducted primarily in youth with polyarticular disease (five or more joints) or with a peripheral joint course affecting a minimum of three joints. In a recent international cohort of youth with axJSpA, only 54% had peripheral arthritis and 11.4% had polyarticular course arthritis.[3] Given the recent advances in novel therapies for adults with SpA, it is imperative that the field is ready to evaluate these targeted therapies in youth, but it is unclear if traditional JIA outcomes, which are primarily focused on measurement of peripheral disease burden, will be responsive and discriminative in this population.

In 2018, the juvenile arthritis Outcome Measures in Rheumatology (OMERACT) working group developed a core domain set prioritized by parents, patients, health care providers, researchers, and regulators.[4] Domains were labeled as "mandatory," "important but optional," and "research agenda" domains. Mandatory domains included joint inflammatory signs, functional limitation, pain, patient assessment of overall well-being, and adverse events. With the OMERACT-recommended framework in mind, the objectives of this study were as follows: (1) to assess the responsiveness, discrimination, and minimal clinically important improvement (MCII) of multiple JIA and SpA outcome measures in a prospective observational cohort of children with axial disease; (2) to assess the potential for subgroup differences to impact these response measures; and (3) to develop a parsimonious list of measures for use in clinical trials of promising new therapies for axJSpA. To accomplish these objectives, we evaluated the measures in a prospective cohort of youth with axial disease treated with TNFi, a therapy used as standard of care and for which efficacy is anticipated for most.

PATIENTS AND METHODS

Participants. The protocol for this prospective study was reviewed and approved by the Institutional Review Board at Children's Hospital of Philadelphia. This was a multicenter prospective longitudinal study of youth with axJSpA evaluated in one of four pediatric hospitals between 2019 and 2024. Potentially eligible youth with JSpA were identified through the staff in the rheumatology clinics at participating centers (Children's Hospital of Philadelphia, Nationwide Children's Hospital, Children's of Alabama, and Cincinnati Children's Hospital Medical Center). Participants were aged 8 to 18 years and met the following inclusion criteria: (1) symptom onset before age 16 years, (2) fulfilled International League of Associations for Rheumatology criteria for enthesitis-related arthritis or psoriatic arthritis[5] or European SpA Study Group criteria,[6] (3) biologic naïve, (4) physician diagnosis of axial arthritis (based on clinical or imaging features), and (5) clinical or imaging features prompting initiation of TNFi therapy.

Assessments. All participants completed the baseline study visit (questionnaires and magnetic resonance imaging [MRI]). Participants without imaging evidence of sacroiliac joint (SIJ) inflammation at baseline completed questionnaires at the 12-week visit. Participants with imaging evidence of SIJ inflammation at baseline completed questionnaires and MRI at the 12-week visit. Participants with continued SIJ inflammation on MRI at 12 weeks completed additional questionnaires and MRI at 24 weeks. Patient-reported outcomes were selected to include the range of domains important to patients and caregivers.[4] Assessments included the following: Patient-Reported Outcomes Measurement Information System (PROMIS) pediatric v2.0 short form pain interference 8a, upper extremity function 8a, mobility 8a, fatigue 10a, and pediatric scale v1.0 global health 7 (each reported as T scores with a mean of 50 and SD of 10, with higher scores representing more of the concept being measured); the Bath Ankylosing Spondylitis Functional Index (BASFI)[7]; the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)[8]; the patient global well-being visual analog scale (VAS) (range 0–10); the patient global disease activity assessment (VAS, range 0–

10); and patient pain intensity (VAS, range 0–10). At the 12- and 24-week visits, participants answered the following question: "Since the start of the study, overall my SpA has: improved, stayed the same, worsened." If the child indicated worsening or improvement, they were asked to rate the magnitude of the change as follows: not important, a little important, moderately important, or extremely important. The PROMIS short forms have been validated in children with juvenile arthritis.[9] The interrater reliability of the BASFI was demonstrated in children and adults with juvenile-onset SpA.[10,11]

Physician-reported metrics collected included the following: swollen joint count, tender joint count, tender enthesis count, and physician global disease activity assessment (VAS, range 0–10). Peripheral joint and tender enthesis examinations were not protocolized but included any involved joint or entheses (except the SIJ) to a maximum of 10, in accordance with how each are quantified for the clinical Juvenile Arthritis Disease Activity Score (cJADAS10) and Juvenile Spondyloarthritis Disease Activity Index (JSpADA).[3]

Overall and composite disease activity metrics collected at all study visits included the Ankylosing Spondylitis Disease Activity Score with C-reactive protein (ASDAS-CRP),[12] BASDAI,[13] JSpADA,[3] and cJADAS.[13] The interrater reliability of the BASDAI has been demonstrated in youth with juvenile-onset SpA.[11] The ASDAS-CRP has not been validated in youth with SpA but has been validated in adults with juvenile-onset SpA.[14] The JSpADA, the original and modified versions, has been validated in youth with SpA.[15,16] The cJADAS has been validated in youth with juvenile arthritis.[13] Binary response outcomes that were assessed included the pediatric the American College of Rheumatology pediatric 50%, 70%, and 90% (ACR Pedi 50, 70, 90) improvement criteria[17]; BASDAI 50% improvement[18]; Assessment of SpondyloArthritis international Society 40% improvement criteria (ASAS40)[19]; ASDAS-CRP clinically important improvement (CII)[20]; ASDAS-CRP major improvement (MI)[21]; and cJADAS10 minimal and inactive disease.[13]

Imaging. MRI included a coronal oblique with STIR, a coronal oblique T1-weighted turbo spin echo, and an axial T2-weighted turbo spin echo with fat saturation. The central imaging team consisted of three raters with extensive experience interpreting pelvic MRI (WPM, NAC, and MLF). MRI reviews of deidentified Digital Imaging and Communications in Medicine images were completed using scoring modules on cearthritis.com and included a detailed assessment for inflammatory lesions in the SIJs using the Spondyloarthritis Research Consortium of Canada (SPARCC) SIJ inflammation score (SIS).[22] The SIS is a validated, objective, responsive, and discriminatory measure for youth with SpA.[23,24] All raters completed recalibration exercises before assessment, and imaging was rated independently and blinded to clinical details by all three central imaging team members at the study end, agnostic to time point.

Analysis. Demographics, clinical features, and measurements collected on the study population were summarized by mean and SD or median and interquartile range (IQR) as appropriate. Age- and sex-specific z scores for weight, height, and body mass index (BMI) were calculated based on the 2000 Centers for Disease Control and Prevention (CDC) growth data for the patient's baseline visit.[25] Overweight status was defined as a BMI ≥85th percentile as per the CDC category definitions for children and teenagers.

Responsiveness to clinical disease activity over time was assessed graphically and by mean change in scores and standardized response mean (SRM) between weeks 0 and 12. SRM was calculated by dividing the mean change between weeks 0 and 12 by the SD of the change. SRM values of 0.2 to 0.5 were considered small, values of 0.5 to 0.8 were considered medium, and values >0.8 were considered large evidence of responsiveness.

Discrimination was assessed in each measure by comparing the median change from the baseline to the 12-week visit between participants who were "responders" and those who were "nonresponders" according to the patient assessment of global well-being using the Wilcoxon rank sum test and standardized mean difference (SMD). SMD was calculated by dividing the mean difference between responders and nonresponders by the pooled SD of those groups. The primary definition for responders was patients with a global well-being assessment that improved ≥2 or patients who had improvement <2 and a value of 0 at week 12; all others were considered nonresponders. Responders were secondarily defined as those who self-reported improvement that was of at least a little importance; all others were defined as nonresponders. SMD values of 0.2 to 0.5 were considered small, values of 0.5 to 0.8 were considered medium, and values >0.8 were considered large discrimination.

MCII. Empirical cut points for the MCII, or the smallest change in measurement that signifies a meaningful improvement in the measure, were determined using an anchor-based (receiver operating characteristic [ROC] curve) method using Stata package 'cutpt.'[26] An empirical anchor-based approach is the primary approach recommended by the US Food and Drug Administration (FDA) for estimation of meaningful change.[27] Patient-reported improvement of at least "a little importance" and change in the patient-reported global assessment of well-being VAS of ≥2 or a follow-up score of 0 were selected as anchors for the anchor-based methods. Change in well-being ≥2 was based on prior work that demonstrated this was a meaningful change.[28,29] Face validity and diagnostic test statistic performance of the different MCIIs were evaluated.

Measure of correlation within domains. Correlation coefficients were used to assess pairwise correlation between each of the measures to identify highly convergent measures within the same domain, adjusted for repeated measures by subject using the R package 'rmcorr.' Within each domain, when measures

were highly correlated (r > |0.8|), those with the largest SMD were prioritized for the final list of measures for the axJSpA core domain.

Subgroup analysis. Differences in the core set measures' SRMs between prespecified subgroups (sex, peripheral arthritis [yes or no], enthesitis [yes or no], and obesity status [BMI z score < or ≥85th percentile]) were determined. We also tested for differences in the proportions meeting binary response criteria (ACR 50/70/90, BASDAI 50% improvement, ASAS40, ASDAS CII and ASDAS[21] MI, cJADAS10 minimal disease, and cJADAS10 inactive disease), stratified by the same subgroups.

Statement of ethics and consent. The protocol for this study was reviewed and approved by a central institutional review board (IRB), the Children's Hospital of Philadelphia's Committee for the Protection of Human Subjects (IRB 19-016713). All participating site IRBs reviewed the application and acknowledged reliance on the central IRB.

RESULTS

The Consolidated Standards of Reporting Trials diagram is shown in Supplementary Figure S1. Seventy-five participants enrolled, 73 (97.3%) of whom completed the baseline visit. Of the participants who completed the baseline visit, 62 (84.9%) had MRI findings consistent with axJSpA, and 57 of 62 (91.9%) completed the 12-week visit. Participant characteristics are shown in Table 1. Of the 57 participants who completed the week 12 visit, 68.4% were male, the median age at enrollment was 15.3 years (IQR 13.5–16.9 years), 41.5% were HLA–B27 positive, 9.3% had a family history of SpA, and 49 met axJSpA classification criteria.[3] The eight participants who failed to meet axJSpA criteria were between 5 and 28 points below the threshold. The most common reasons participants did not reach any level in a domain and received a score of zero were not meeting the unequivocal structural (n = 7) or inflammatory (n = 5) lesion definitions, no relevant family history in first-degree relative, and HLA–B27 unknown or negative (n = 6). The axJSpA classification criteria were finalized and published after enrollment for this study was complete and were available for use as inclusion criteria. Forty-nine (86.0%) patients were treated with adalimumab, 5 (8.8%) were treated with etanercept, 2 (3.5%) were treated with golimumab, and 1 (1.8%) was treated with infliximab. Of youth with SpA and sacroiliitis, 82.7% reported a clinical improvement of "moderate" or "large" importance to TNFi therapy by 12 weeks. Thirty-three (58%) participants had continued MRI inflammation at 12 weeks, 76% of which reported clinical improvement of at least moderate importance (patient-reported clinical status unavailable on three participants). Twenty-five (75.8%) of 33 participants with continued inflammation on MRI at 12 weeks completed the 24-week study visit. Only 12.3% of this cohort would meet current JIA trial entry criteria necessitating three or more active peripheral joints.

Responsiveness. Table 2 shows the change scores and responsiveness as measured by the SRM for continuous measures from weeks 0 to 12, grouped by domains prioritized by the JIA OMERACT working group. Within the pain domain, PROMIS pain interference (0.77) and the patient pain intensity VAS (0.77) were most responsive. In the physical function domain, PROMIS mobility (0.83) and the BASFI (0.81) were most responsive. Within the joint inflammatory signs domain, the tender enthesis count (0.54) and the MRI SPARCC SIS (1.02) were most responsive. Among measures of patient perception of well-being, the patient global well-being VAS (0.88) and patient disease activity VAS (0.72) were most responsive. All the overall and composite disease activity response measures were responsive, but the JSpADA8 (1.28) and cJADAS10 were the largest (1.47). Box plots of change scores between responders and nonresponders are shown in Supplementary Figure S2.

Thirty-seven (68.5%), 33 (61.1%), 23 (46%), and 15 (30.6%) participants met the pediatric ACR30/50/70/90 outcomes. Twenty (35.1%) and 32 (56.1%) met the cJADAS10 inactive disease and minimal disease outcomes, respectively. Twenty-one (37.5%), 18 (50%), 4 (11.1%), and 26 (46.4%) met ASAS40, ASDAS-CRP CII, ASDAS MI, and BASDAI 50% improvement, respectively.

Discrimination. Forty participants were classified as treatment responders, and 17 participants were nonresponders. Discrimination, as measured by the SMD, for all measures is shown in Figure 1. Within the pain domain, the axial pain intensity VAS (0.67) and PROMIS pain interference (0.50) were the best discriminatory measures. In the physical function domain, PROMIS mobility (0.75) was most discriminatory, and the remainder of measures had small discrimination. Within the joint inflammatory signs domain, only the MRI SPARCC SIS (0.53) had at least medium discrimination. Among measures of patient perception of well-being, PROMIS global health (0.52) had medium discrimination; the patient global well-being VAS SMD was not measured against the primary responder definition because it was used to define responder status but had large SMD against the secondary responder definition (Supplementary Figure S3). Of the overall and composite disease activity measures, the JSpADA8 (0.89) and the ASDAS (0.90) had large discrimination. Both seven-component versions of the JSpADA and the BASDAI had medium discrimination. The cJADAS10 SMD was not measured because patient global assessment, which was used to define responder status, is a component of the score. The pediatric ACR30/50/70/90, cJADAS10 inactive disease and minimal disease, and BASDAI 50% improvement all discriminated between responders and nonresponders (all $P < 0.01$).

Table 1. Baseline participant characteristics*

	All participants		Participants with MRI inflammation at baseline		Participants with ≥2 MRI scans	
	n	Median (IQR) or n (%)	n	Median (IQR) or n (%)	n	Median (IQR) or n (%)
Criteria fulfilled						
ESSG	73	55 (75.3)	62	51 (82.3)	57	48 (84.2)
ILAR ERA	73	60 (82.2)	62	53 (85.5)	57	49 (86.0)
ILAR PsA	73	4 (5.5)	62	3 (4.8)	57	3 (5.3)
axJSpA	73	48 (65.8)	62	48 (77.4)	57	48 (84.2)
Age at baseline	73	15.5 (13.6–17.0)	62	15.4 (13.5–17.0)	57	15.3 (13.5–16.9)
Sex, male	73	45 (61.6)	62	41 (66.1)	57	39 (68.4)
HLA–B27 positive	66	30 (45.5)	57	25 (43.9)	53	22 (41.5)
Family history of SpA	68	9 (13.2)	58	7 (12.1)	54	5 (9.3)
BMI category, obesity	73	8 (11.0)	62	6 (9.7)	57	5 (8.8)
Polyarticular course	73	7 (9.6)	62	4 (6.5)	57	4 (7.0)
AJC ≥3	73	11 (15.1)	62	7 (11.3)	57	7 (12.3)
BMI ≥85th percentile	73	21 (28.8)	62	16 (25.8)	57	14 (24.6)
Peripheral arthritis	73	20 (27.4)	62	16 (25.8)	57	16 (28.1)
Enthesitis	73	39 (53.4)	62	33 (53.2)	57	31 (54.4)
Psoriasis	73	6 (8.2)	62	6 (9.7)	57	6 (10.5)
Inflammatory bowel disease	73	9 (12.3)	62	9 (14.5)	57	9 (15.8)
BASDAI[a] (0–10)	72	4.7 (3.5–6.2)	61	4.5 (3.3–6.1)	56	4.4 (3.1–6.2)
JSpADA8[a,b] (0–8)	49	3.5 (3.0–4.5)	42	3.5 (3.0–4.5)	38	3.5 (3.0–4.5)
ASDAS–CRP[a] (0 to no upper limit)	65	2.9 (2.2–3.4)	55	2.8 (2.2–3.2)	51	2.7 (2.2–3.2)
cJADAS10[a] (0–30)	72	9.0 (7.0–11.0)	61	9.0 (6.0–11.0)	56	9.0 (6.0–11.0)
Physician global assessment (0–10)	73	3.0 (2.0–4.0)	62	3.0 (2.0–4.0)	57	3.0 (2.0–4.0)
SPARCC sacroiliac joint inflammation score (0–72)[c]	–	–	–	–	57	11.0 (5.0–20.0)

* AJC, active peripheral joint count; ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score with C-reactive protein; axJSpA, axial juvenile spondyloarthritis; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BMI, body mass index; cJADAS10, clinical Juvenile Arthritis Disease Activity Score; ERA, enthesitis-related arthritis; ESSG, European Spondyloarthropathy Study Group; ILAR, International League of Associations for Rheumatology; IQR, interquartile range; JSpADA8, 8-item (full) Juvenile Spondyloarthritis Disease Activity Index; MRI, magnetic resonance imaging; PsA, psoriatic arthritis; SpA, spondyloarthritis; SPARCC, Spondyloarthritis Research Consortium of Canada.
[a] Higher scores indicate more active disease.
[b] The JSpADA index consists of 8 equally weighted measures.
[c] Only participants with ≥2 MRI scans were reviewed by the central imaging team.

Correlation within domains and development of parsimonious core set. Within the pain domain, the patient pain VAS and axial pain VAS (r = 0.73) and axial pain and PROMIS pain interference (r = 0.67) had high correlation. Of these three measures, only PROMIS pain interference had a large SRM and SMD and was chosen for the core set.

In the physical function domain, the BASFI and PROMIS mobility were highly correlated (r = 0.73) and of these, only PROMIS mobility had large responsiveness and discrimination. PROMIS upper extremity had medium responsiveness but low discrimination. The PROMIS mobility measure was chosen for the core set.

Within the joint inflammatory signs domain, all the measures had low correlation, with active and tender joint counts having the highest in the domain (r = 0.45). The only measure with large responsiveness and at least medium discrimination was the SPARCC SIS. The core set for this domain includes the SPARCC SIS.

Among measures of patient perception of well-being, patient global assessments of well-being and disease activity were moderately correlated (r = 0.68). Both had large responsiveness, but only patient global assessment of well-being was discriminative according to the secondary responder definition (Supplementary Figure S3). Of these measures, only the patient assessment of well-being VAS was chosen for the core set.

All the overall and composite disease activity measures had moderate to high correlation (all except physician global assessment had r > 0.7; physician global assessment and the other measures had moderate correlation of r > 0.54). All these disease measures had high responsiveness (≥0.8) and high discrimination (≥0.8), except the physician global assessment, which had small discrimination. Collection of at least one of the overall and composite measures is recommended. The recommended core set of measures for axJSpA trials is listed in Table 3.

MCII. Table 4 lists the MCII for each core set measure. The anchor-based methods using ROC analysis had the optimum combination of face validity and test statistic performance. The area under the ROC curve (AUROC) for all measures except for mobility (0.67) and SIS (0.68) was greater than 0.70. Of the overall and composite disease activity measures, the JSpADA8 had the highest AUROC, sensitivity, and specificity.

Table 2. Change scores and SRM*

Measures	Change from index to 12 wk		
	n	Median (IQR)	SRM
Pain[a]			
PROMIS pain interference	56	−7.6 (−16.0 to 0.0)	0.77
Patient pain intensity VAS	56	−2.0 (−4.5 to 0.0)	0.77
Patient assessment of neck, back, and hip pain	56	−1.5 (−4.5 to 0.0)	0.62
Joint inflammatory signs[a]			
Active joint count	57	0.0 (0.0 to 0.0)	0.29
Tender joint count	57	0.0 (−1.0 to 0.0)	0.40
Tender enthesis count	57	0.0 (−2.0 to 0.0)	0.54
SPARCC sacroiliac joint inflammation score	57	−8.0 (−18.0 to −3.0)	1.02
Activity limitation/physical function[a]			
PROMIS upper extremity function	56	−0.0 (−0.0 to 9.3)	0.47
PROMIS mobility	56	6.1 (0.0 to 15.9)	0.83
BASFI	56	−1.1 (−2.8 to 0.0)	0.81
Patient perception of disease (well-being)[a]			
Patient global well-being VAS	56	−2.0 (−4.0 to −1.0)	0.88
Patient global disease activity VAS	54	−2.0 (−4.0 to 0.0)	0.72
PROMIS fatigue	56	−6.6 (−12.8 to 1.7)	0.41
PROMIS global health	56	5.1 (−0.1 to 10.3)	0.64
Overall/composite response measures			
BASDAI	56	−1.6 (−3.3 to 0.0)	0.85
JSpADA8	32	−2.0 (−3.0 to −1.0)	1.28
JSpADA7, no markers of inflammation	36	−1.5 (−2.5 to −0.5)	1.13
JSpADA7, no modified Schober's test	49	−1.5 (−2.5 to −1.0)	1.14
JSpADA6	56	−1.5 (−2.0 to −0.5)	1.05
ASDAS-CRP	34	−1.2 (−1.9 to −0.2)	0.83
cJADAS10	56	−5.0 (−7.0 to −3.0)	1.47
Physician global assessment VAS	57	−2 (−3 to −1)	1.29

* SRM values of 0.2–0.5 were considered small, values of 0.5–0.8 were considered medium, and values >0.8 were considered large evidence of responsiveness. All raw scores generated from PROMIS instruments are translated into standardized T scores with a population mean of 50 and SD of 10. Higher scores in a domain represent more of the trait being measured; higher T scores indicate a worse outcome in the following domains: fatigue and pain interference; lower T scores indicate a worse outcome in the remaining domains. The BASDAI range is 0–10, with higher scores indicating more active disease. The BASFI range is 0–10, with higher scores indicating more impairment. All VAS ranges were 0–10, with higher scores indicating a worse state. The JSpADA index consists of 8 equally weighted measures and is scored 0–8, with higher scores indicating more active disease; the two 7-component variants both have a score range of 0–7. ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score with C-reactive protein; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; cJADAS10, clinical Juvenile Arthritis Disease Activity Score with maximum 10 active joint count; IQR, interquartile range; JSpADA6, 6-item Juvenile Spondyloarthritis Disease Activity Index; JSpADA7, 7-item Juvenile Spondyloarthritis Disease Activity Index; JSpADA8, 8-item (full) Juvenile Spondyloarthritis Disease Activity Index; PROMIS, Patient-Reported Outcomes Measurement Information System; SPARCC, Spondyloarthritis Research Consortium of Canada; SRM, standardized response mean; VAS, visual analog score.
[a] Mandatory domain per the juvenile idiopathic arthritis Outcome Measures in Rheumatology working group.

Subgroup analysis. Responsiveness, as determined by the SRM, for prespecified subgroups for each core set of measure is shown in Figure 2. The largest differences were seen in those who were overweight (BMI ≥85th percentile) versus those with normal weight. Overweight youth had lower responsiveness than youth with normal weight for the pain, function, joint inflammatory signs and perception of disease measures and outcomes, with the most notable differences for PROMIS pain interference (0.04 vs 1.04), PROMIS mobility (0.50 vs 0.93), patient global well-being (0.60 vs 0.97), and BASDAI (0.40 vs 1.05). Responsiveness was higher in overweight individuals for most of the overall and composite disease activity measures, most prominently for JSpADA8 (1.89 vs 1.18) and cJADAS10 (1.87 vs 1.37).

Male participants had higher responsiveness for the MRI SIS (1.12 vs 0.88) than female participants but also had higher baseline mean SIS scores (15.3 vs 10.2). Responsiveness of patient perception of disease was higher for female participants (1.28 vs 0.78). In male and female participants, responsiveness for overall disease measures and measures of pain and function were similar.

In the evaluation of those with enthesitis versus those without enthesitis, differences in responsiveness were most notable for the SPARCC SIS (0.51 vs 0.31). Responsiveness for overall disease measures weas higher in youth with enthesitis, except for the ASDAS-CRP and cJADAS10.

In subgroup responsiveness analysis of those with and without peripheral arthritis, youth with arthritis had lower PROMIS pain

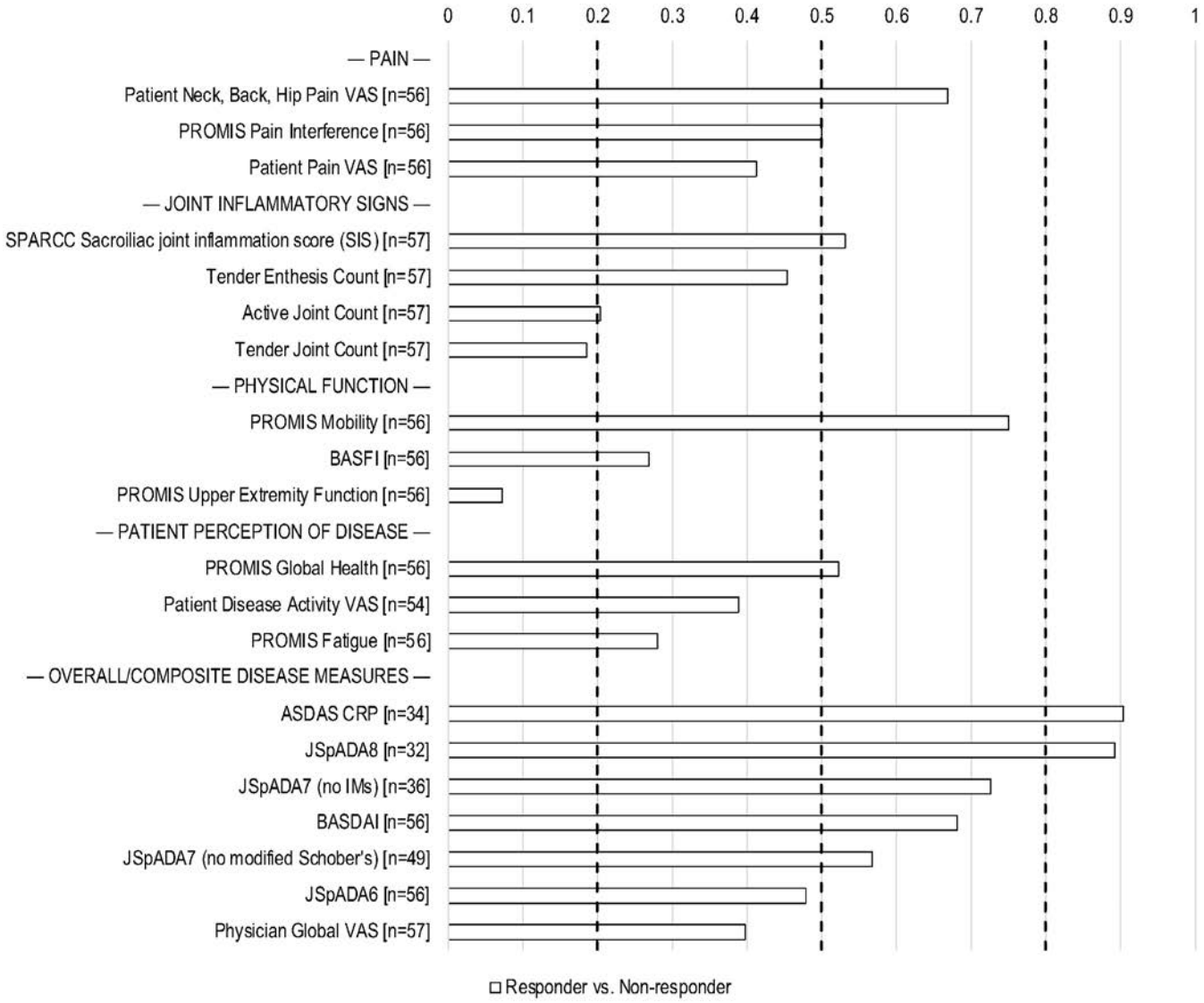


Figure 1. Standardized mean difference at the 12-week visit by clinical responder status. "Clinical responder" is defined as patient global assessment VAS improvement of ≥2 or a score of 0 at week 12. Standardized mean difference values of 0.2 to 0.5 were considered small, values of 0.5 to 0.8 were considered medium, and values >0.8 were considered large evidence of discrimination. ASDAS CRP, Ankylosing Spondylitis Disease Activity Score with C-reactive protein; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; BASFI, Bath Ankylosing Spondylitis Functional Index; IMs, markers of inflammation; JSpADA7, 7-item Juvenile Spondyloarthritis Disease Activity Index; JSpADA8, 8-item (full) Juvenile Spondyloarthritis Disease Activity Index; PROMIS, Patient-Reported Outcomes Measurement Information System; SPARCC, Spondyloarthritis Research Consortium of Canada; VAS, visual analog score.

interference (0.39 vs 0.98) and patient global well-being VAS (0.74 vs 0.93), and the majority of overall disease activity measures were higher (all except cJADAS10 and BASDAI). There were no differences in responsiveness for PROMIS mobility for those with and without peripheral arthritis.

DISCUSSION

In this prospective real-world axJSpA population treated with TNFi, the most responsive measures over 12 weeks were the SPARCC SIS, PROMIS pain interference, the patient global well-being VAS, and the physician global disease activity assessment. Several important key findings are worth highlighting. First, most youth with sacroiliitis, as expected, had a clinical and imaging response to TNFi therapy. Of youth with SpA and sacroiliitis, 82.7% reported a clinical improvement of "moderate" or "large" importance to TNFi therapy by 12 weeks. Of youth treated with TNFi, 70.2% met the responder definition by 12 weeks, and although 58% (33 of 57) of participants had continued MRI inflammation at 12 weeks, 77% of those patients reported clinical

Table 3. Core set of outcome measures for axial juvenile spondyloarthritis*

Domain	Measure
Pain	PROMIS pain interference
Activity limitation/physical function	PROMIS mobility
Joint inflammatory signs	SPARCC SIS
Patient perception of disease (well-being)	Patient assessment of well-being VAS
Overall/composite response measures	Choose 1 of the following: JSpADA (v8 or either v7 version), cJADAS10, ASDAS-CRP, BASDAI
Adverse events	As per investigator or funding requirement
Strongly suggested, but optional	Tender entheses count, SPARCC SIJ structural score, stiffness ≥15 min, measure of serologic inflammation

* Mandatory juvenile idiopathic arthritis domains determined based on those identified by the juvenile arthritis Outcome Measures in Rheumatology working group.[4] ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score with C-reactive protein; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; cJADAS10, clinical Juvenile Arthritis Disease Activity Score with maximum 10 active joint count; JSpADA, Juvenile Spondyloarthritis Disease Activity Index; PROMIS, Patient-Reported Outcomes Measurement Information System; SIJ, sacroiliac joint; SIS, sacroiliac joint inflammation score; SPARCC, Spondyloarthritis Research Consortium of Canada; VAS, visual analog score.

improvement of at least moderate importance. Second, peripheral disease (arthritis or enthesitis) occurs in only half of patients. Third, more traditional measures used to assess medication efficacy, such as active joint count and physician global disease assessment, did not perform well in this population and had either low responsiveness, low discrimination, or both. Fourth, only 12.3% of the cohort presented herein would be eligible for enrollment in a JIA trial using current inclusion criteria, which necessitate at least three active peripheral joints. Fifth, there were important subgroup differences in responsiveness, most notably in youth who were overweight versus those with normal weight. These responsiveness differences were observed for pain, mobility, and well-being. Lastly and most importantly, there is now a parsimonious set of measures that can assess the novel efficacy of medications for this understudied condition.

The OMERACT JIA working group concluded the following domains were mandatory in JIA trials: joint inflammatory signs, functional limitation, pain, patient assessment of overall well-being, and adverse events. In JIA trials (particularly of the polyarticular course), the joint inflammatory signs domain most commonly assesses active joint and tender joint counts. When considering JSpA, one might also consider the tender enthesis count. However, none of these three counts were responsive or discriminative in the population with axJSpA. Similar findings of low responsiveness of tender and swollen joint counts have been reported in adults with SpA, namely those with psoriatic arthritis (PsA).[30] Youth with axJSpA, as evidenced by this cohort and others, often do not have a large (if any) burden of peripheral disease, making the lack of responsiveness not entirely surprising. The SIJs, additionally, are tricky to reliably examine because the joint does not have detectable swelling, and tenderness on examination is often nonspecific.[31] The only measure that was both reliable and discriminative in this population for joint inflammatory signs was the MRI SPARCC inflammation score. In fact, the lack

Table 4. Cutoffs and performance of anchor-based MCII values in select outcome measures for patients with axial juvenile spondyloarthritis*

	MCII	AUROC	Sensitivity	Specificity
Pain				
PROMIS pain interference	−3.2	0.88	1.0	0.75
Physical function				
PROMIS mobility	3.5	0.67	0.65	0.69
Joint inflammatory signs				
SPARCC sacroiliac joint inflammation score	−5.5	0.68	0.71	0.64
Patient perception of disease				
Patient global VAS	−2.5	0.78	1.0	0.57
Overall/composite disease measures				
JSpADA8	−0.8	0.94	1.00	0.88
JSpADA7, no markers of inflammation	−0.8	0.80	0.80	0.79
JSpADA7 no modified Schober's test	−1.3	0.86	1.0	0.71
cJADAS10	−3.5	0.80	0.86	0.75
BASDAI	−1.8	0.78	1.0	0.57
ASDAS-CRP	−0.5	0.78	0.83	0.72

* ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score with C-reactive protein; AUROC, area under the receiver operating characteristic curve; BASDAI, Bath Ankylosing Spondylitis Disease Activity Index; cJADAS10, clinical Juvenile Arthritis Disease Activity Score with maximum 10 active joint count; JSpADA7, 7-item Juvenile Spondyloarthritis Disease Activity Index; JSpADA8, 8-item (full) Juvenile Spondyloarthritis Disease Activity Index; MCII, minimal clinically important improvement; PROMIS, Patient-Reported Outcomes Measurement Information System; SPARCC, Spondyloarthritis Research Consortium of Canada; VAS, visual analog score.

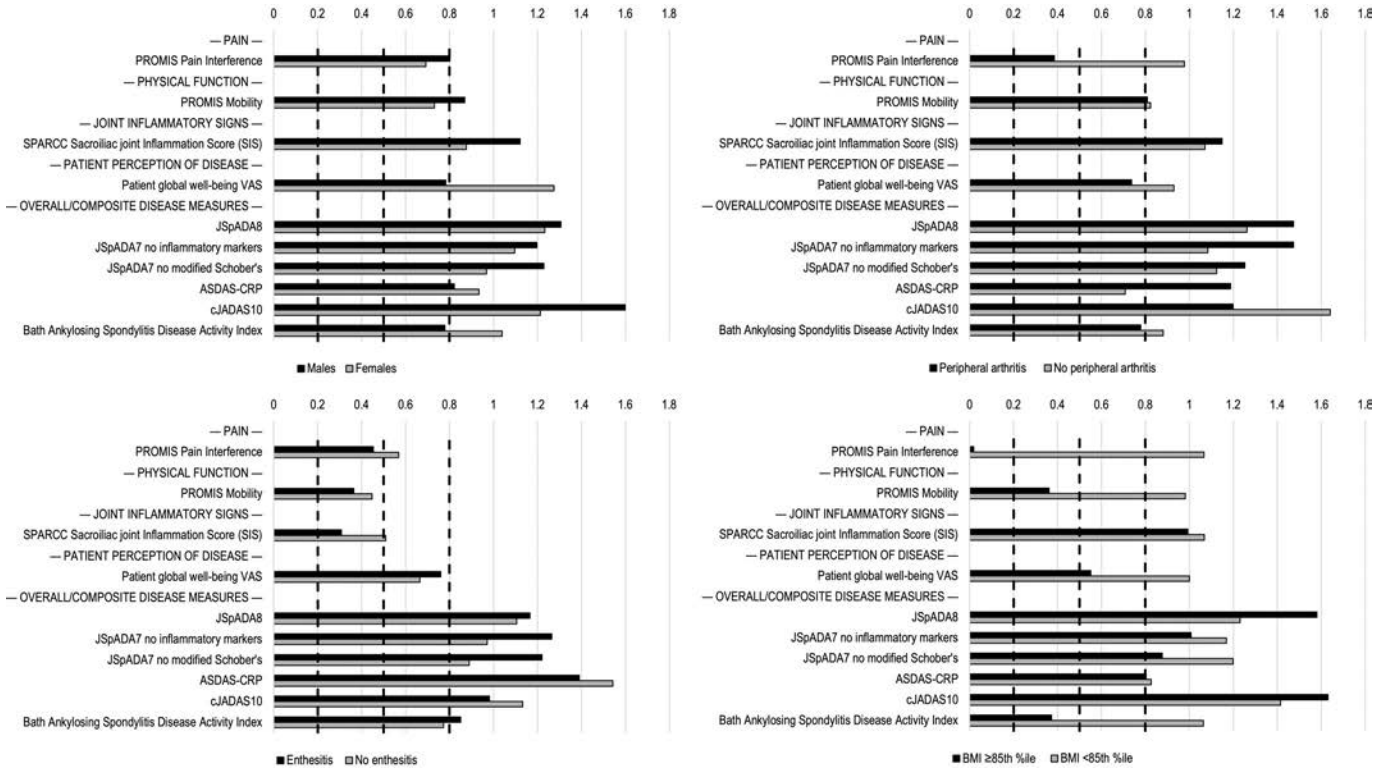


Figure 2. Standardized response mean by subgroup. Subgroups analyzed were the following: (top left) biologic sex (males vs females); (top right) peripheral arthritis (peripheral arthritis vs no peripheral arthritis at baseline); (bottom left) enthesitis (enthesitis vs no enthesitis at baseline), and (bottom right) BMI (≥85th percentile vs <85th percentile). The numbers of patients for each subgroup analysis are as follows: sex: male = 39, female = 18; enthesitis: yes = 31, no = 26; peripheral arthritis: yes = 16, no = 41; BMI 85th percentile: yes = 14, no = 43. ASDAS-CRP, Ankylosing Spondylitis Disease Activity Score with C-reactive protein; BMI, body mass index; cJADAS10, clinical Juvenile Arthritis Disease Activity Score with maximum 10 active joint count; JSpADA7, 7-item Juvenile Spondyloarthritis Disease Activity Index; JSpADA8, 8-item (full) Juvenile Spondyloarthritis Disease Activity Index; PROMIS, Patient-Reported Outcomes Measurement Information System; SPARCC, Spondyloarthritis Research Consortium of Canada; VAS, visual analog score.

of imaging to assess TNFi efficacy at the SIJs in the earlier published trial of youth with axial disease may explain the ultimately negative result.[1] Despite the costs of MRI, inclusion of imaging for evaluation of novel therapeutics' effectiveness at the SIJ is imperative to carefully consider. Correlation of the MRI inflammation score and the measure used to define responder status, patient global assessment, was only moderate (0.56) because they are both measuring different but equally important concepts. In fact, the MRI inflammation score was not highly correlated with any of the other measures tested, again underscoring its importance in assessing the effectiveness of the drug at its intended target. Without imaging, dependence on measures that heavily rely on the presence of peripheral disease may result in a false-negative conclusion. Regarding the other three mandatory OMERACT domains of functional limitation, pain, and patient assessment of overall well-being, there were easy-to-administer measures that demonstrated responsiveness and discrimination for this population.

All the overall and composite disease activity measures tested, except the physician global assessment, were responsive and discriminatory. Ultimately, the choice of what, if any, composite measure to use should take into consideration the burden of peripheral disease activity at baseline in the target population, whether the drug under study is anticipated to have differential effects of certain disease manifestations (ie, peripheral arthritis, axial arthritis, dactylitis, or enthesitis), whether the trial is intended to fulfill FDA or European Medicines Agency (EMA) labeling requirements, and/or whether the measures have been validated in youth. Of the tested measures, the pediatric ACR30 is recognized as the gold standard for assessing treatment response in polyarticular course JIA and is accepted by the FDA and EMA for drug registration.[17,32] The pediatric ACR core set, cJADAS10, and JSpADA are all validated measures in youth.

Subgroup analysis demonstrated some interesting and important differences. In the subgroup analysis of youth with overweight versus those without overweight, there was a consistent and blunted response for nearly all measures among overweight individuals but most notably for PROMIS pain interference, PROMIS mobility, and global well-being. Additionally, overweight youth started with lower or better baseline scores

for all measures in comparison to normal weight youth. Differences in therapy response based on obesity status has been shown in adults with arthritis,[33] and a negative association between BMI percentile and TNFi concentration has been shown in youth with JIA.[34] In a meta-analysis of select inflammatory diseases, including rheumatoid arthritis, SpA, and PsA, adults with obesity had a 60% higher odds of TNFi response failure, and the response was dose dependent, with increased odds of 6.5% with each 1-kg/m^2 increase in BMI.[33] This altered response to TNFi is likely multifactorial and related to higher systemic inflammatory burden[35,36] as well as altered drug pharmacokinetics, including increased clearance, lower absorption, and alterations in the volume of distribution.[37,38] Additionally, the SRM for the MRI SIS was larger in male participants, and this may be at least partially explained by higher baseline mean SIS scores (15.3 vs 10.2). In adults with ankylosing spondylitis and nonradiographic spondylitis, sex differences have been reported frequently and include more SIJ inflammation, higher prevalence of fat lesions, and more structural damage in men compared to women.[39,40] Given these findings, consideration should be given to stratification of analysis based on BMI (≥85th percentile vs <85th percentile) if obesity is prevalent in the study population and possibly also to biologic sex if imaging is a primary outcome.

Strengths of this study include the multicenter cohort, evaluation of a comprehensive set of pediatric JIA and adult SpA outcomes, blinded central imaging scoring, and testing the responsiveness and discrimination of these measures in a real-world setting. However, several limitations should be considered. As with any prospective study, there were some missing data, albeit minimal for all measures. Some physical examination metrics and adult SpA instruments were added to the patient questionnaires later than others, resulting in different numbers of patients for several of the adult outcomes. Evaluation for axial disease in this cohort was done per the treating physician. Because thresholds to initiate this evaluation may differ across centers and by provider, asymptomatic and mild cases may have been missed. However, those with milder cases are unlikely to be the patients who would be considered for initiation of biologic therapy and/or candidates for a pragmatic therapy trial for axial disease.

In conclusion, most youth with axJSpA had a clinical and imaging response to TNFi therapy. Some of the most responsive and discriminative measures in this population were not the standard outcomes for current polyarticular course JIA trials. Moreover, this study's results suggest that standard calculated response metrics used in pediatric clinical trials may underestimate clinical response in children with axial disease, and subgroup analysis demonstrated some interesting differences regarding overweight status and biologic sex that warrant consideration in the design of future trials. Lastly, and most importantly, there is now a parsimonious set of measures that can assess the efficacy of novel medications for this understudied condition.

ACKNOWLEDGMENTS

The authors thank Joel Paschke (CARE Arthritis) for preparing the imaging review modules for web-based assessment and data entry and the following collaborators for their contributions in patient recruitment, study procedure scheduling, and data collection and verification: Livie Timmerman (University of Alabama-Birmingham), Joanne Drew (Nationwide Children's Hospital), Madelyn Jones (Nationwide Children's Hospital), Megan Quinlan-Waters (Cincinnati Children's Hospital Medical Center), Colin Burns (Children's Hospital of Philadelphia), and Cassandra Muir (Children's Hospital of Philadelphia).

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 Weiss 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. Horneff G, Fitter S, Foeldvari I, et al. Double-blind, placebo-controlled randomized trial with adalimumab for treatment of juvenile onset ankylosing spondylitis (JoAS): significant short term improvement. Arthritis Res Ther 2012;14(5):R230.
2. 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(6):717–734.
3. Weiss PF, Brandon TG, Aggarwal A, et al. Classification criteria for axial disease in youth with juvenile spondyloarthritis. Arthritis Rheumatol 2024;76(12):1797–1808.
4. Morgan EM, Munro JE, Horonjeff J, et al. Establishing an updated core domain set for studies in juvenile idiopathic arthritis: a report from the OMERACT 2018 JIA workshop. J Rheumatol 2019;46(8):1006–1013.
5. Petty RE, Southwood TR, Manners P, et al; International League of Associations for Rheumatology. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31(2):390–392.
6. Dougados M, van der Linden S, Juhlin R, et al. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34(10):1218–1227.
7. Calin A, Garrett S, Whitelock H, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol 1994;21(12):2281–2285.
8. Garrett S, Jenkinson T, Kennedy LG, et al. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994;21(12):2286–2291.
9. Brandon TG, Becker BD, Bevans KB, et al. Patient-Reported Outcomes Measurement Information System tools for collecting patient-reported outcomes in children with juvenile arthritis. Arthritis Care Res (Hoboken) 2017;69(3):393–402.
10. Wei JC, Wong RH, Huang JH, et al. Evaluation of internal consistency and re-test reliability of Bath ankylosing spondylitis indices in a large

cohort of adult and juvenile spondylitis patients in Taiwan. Clin Rheumatol 2007;26(10):1685–1691.

11. Batthish M, Rachlis A, Wong B, et al. Intrarater reliability of the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) and the Bath Ankylosing Spondylitis Functional Index (BASFI) in children with spondyloarthritis [abstract]. Pediatr Rheumatol Online J; 2012:10-(suppl 1):A45.
12. Machado PM, Landewé R, Heijde DV; Assessment of SpondyloArthritis international Society (ASAS). Ankylosing Spondylitis Disease Activity Score (ASDAS): 2018 update of the nomenclature for disease activity states. Ann Rheum Dis 2018;77(10):1539–1540.
13. Trincianti C, Van Dijkhuizen EHP, Alongi A, et al; Paediatric Rheumatology International Trials Organisation. Definition and validation of the American College of Rheumatology 2021 juvenile arthritis disease activity score cutoffs for disease activity states in juvenile idiopathic arthritis. Arthritis Rheumatol 2021;73(11):1966–1975.
14. Sundaram TG, Muhammed H, Aggarwal A, et al. A prospective study of novel disease activity indices for ankylosing spondylitis. Rheumatol Int 2020;40(11):1843–1849.
15. Weiss PF, Colbert RA, Xiao R, et al. Development and retrospective validation of the Juvenile Spondyloarthritis Disease Activity Index. Arthritis Care Res (Hoboken) 2014;66(12):1775–1782.
16. Srinivasalu H, Treemarcki EB, Rumsey DG, et al; CARRA Spondyloarthritis Workgroup and the CARRA Registry Investigators. Modified Juvenile Spondyloarthritis Disease Activity Index in the Childhood Arthritis and Rheumatology Research Alliance (CARRA) Registry. J Rheumatol 2023;50(4):532–537.
17. Giannini EH, Ruperto N, Ravelli A, et al. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40(7):1202–1209.
18. Rudwaleit M, Listing J, Brandt J, et al. Prediction of a major clinical response (BASDAI 50) to tumour necrosis factor α blockers in ankylosing spondylitis. Ann Rheum Dis 2004;63(6):665–670.
19. Brandt J, Listing J, Sieper J, et al. Development and preselection of criteria for short term improvement after anti-TNFα treatment in ankylosing spondylitis. Ann Rheum Dis 2004;63(11):1438–1444.
20. Machado P, Landewé R, Lie E, et al. Ankylosing Spondylitis Disease Activity Score (ASDAS): defining cut-off values for disease activity states and improvement scores. Ann Rheum Dis 2011;70(1):47–53.
21. van der Heijde D, Joshi A, Pangan AL, et al. ASAS40 and ASDAS clinical responses in the ABILITY-1 clinical trial translate to meaningful improvements in physical function, health-related quality of life and work productivity in patients with non-radiographic axial spondyloarthritis. Rheumatology (Oxford) 2016;55(1):80–88.
22. Maksymowych WP, Lambert RG, Brown LS, et al. Defining the minimally important change for the SpondyloArthritis Research Consortium of Canada spine and sacroiliac joint magnetic resonance imaging indices for ankylosing spondylitis. J Rheumatol 2012;39(8):1666–1674.
23. Weiss PF, Maksymowych WP, Xiao R, et al. Spondyloarthritis Research Consortium of Canada sacroiliac joint inflammation and structural scores: change score reliability and recalibration utility in children. Arthritis Res Ther 2020;22(1):58.
24. Weiss PF, Maksymowych WP, Lambert RG, et al. Feasibility and reliability of the Spondyloarthritis Research Consortium of Canada sacroiliac joint inflammation score in children. Arthritis Res Ther 2018;20(1):56.
25. Ogden CL, Kuczmarski RJ, Flegal KM, et al. Centers for Disease Control and Prevention 2000 growth charts for the United States: improvements to the 1977 National Center for Health Statistics version. Pediatrics 2002;109(1):45–60.
26. Clayton P. CUTPT: Stata Module for Empirical Estimation of Cutpoint for a Diagnostic Test. Statistical Software Components; 2013.
27. McLeod LD, Coon CD, Martin SA, et al. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. Expert Rev Pharmacoecon Outcomes Res 2011;11(2):163–169.
28. Weiss PF, Brandon TG, Ryan ME, et al. Development and validation of a juvenile spondyloarthritis disease flare measure: ascertaining flare in patients with inactive disease. Arthritis Care Res (Hoboken) 2023;75(2):373–380.
29. Brunner HI, Lovell DJ, Finck BK, et al. Preliminary definition of disease flare in juvenile rheumatoid arthritis. J Rheumatol 2002;29(5):1058–1064.
30. Karmacharya P, Stull C, Stephens-Shields A, et al. Responsiveness and minimum clinically important difference in patient-reported outcome measures among patients with psoriatic arthritis: a prospective cohort study. Arthritis Care Res (Hoboken) 2023;75(10):2182–2189.
31. Weiss PF, Xiao R, Biko DM, et al. Assessment of sacroiliitis at diagnosis of juvenile spondyloarthritis by radiography, magnetic resonance imaging, and clinical examination. Arthritis Care Res (Hoboken) 2016;68(2):187–194.
32. Balay-Dustrude E, Shenoi S. Current validated clinical and patient reported disease outcome measures in juvenile idiopathic arthritis. Open Access Rheumatol 2023;15:189–206.
33. Singh S, Facciorusso A, Singh AG, et al. Obesity and response to anti-tumor necrosis factor-α agents in patients with select immune-mediated inflammatory diseases: a systematic review and meta-analysis. PLoS One 2018;13(5):e0195123.
34. Balevic SJ, Becker ML, Gonzalez D, et al. Low etanercept concentrations in children with obesity and juvenile idiopathic arthritis. J Pediatr Pharmacol Ther 2021;26(8):809–814.
35. Galic S, Oakhill JS, Steinberg GR. Adipose tissue as an endocrine organ. Mol Cell Endocrinol 2010;316(2):129–139.
36. Lumeng CN, Saltiel AR. Inflammatory links between obesity and metabolic disease. J Clin Invest 2011;121(6):2111–2117.
37. Mould DR. The pharmacokinetics of biologics: a primer. Dig Dis 2015;33(suppl 1):61–69.
38. Lee H, Kimko HC, Rogge M, et al. Population pharmacokinetic and pharmacodynamic modeling of etanercept using logistic regression analysis. Clin Pharmacol Ther 2003;73(4):348–365.
39. Pereira GC, Machado NP, Gomes AF, et al. Sacroiliac and spine imaging in spondyloarthritis: does phenotype or sex matter? Adv Rheumatol 2024;64(1):68.
40. Lorenzin M, Cozzi G, Scagnellato L, et al. Relationship between sex and clinical and imaging features of early axial spondyloarthritis: results from a 48 month follow-up (Italian arm of the SPondyloArthritis Caught Early (SPACE) study). Scand J Rheumatol 2023;52(5):519–529.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1240–1245
DOI 10.1002/acr.25562

Improving Documentation Rates of Contraception and Reproductive Planning in Patients With Rheumatic Disease

Benjamin Moon,[1] Puneet Bajaj,[2] Megan Clowse,[3] James Roberts,[2] Kathryn Dao,[2] Bonnie L. Bermas,[2] and Brooke S. Mills[2]

Objective. This study aimed to improve contraception and reproductive planning documentation within rheumatology providers' notes at a single academic center.

Methods. Female patients aged 18 to 45 years with autoimmune inflammatory rheumatic diseases were identified, and chart review was performed for documentation of contraception and pregnancy planning. Baseline data were collected from 148 charts between May 2022 and March 2023. In June 2023, a reproductive Health Assessment Questionnaire was integrated into the electronic health record and sent to patients for completion before their visits. Postintervention data were collected from 176 charts between July 2023 and December 2023. Demographics of patients (race, ethnicity, and sex) and provider sex were collected. Telehealth and face-to-face visits were assessed separately.

Results. A statistically significant increase ($P < 0.0001$) was seen in provider documentation of both contraception (from 44.6 to 70.5%) and pregnancy planning (from15.5 to 60.2%) after implementation of the previsit questionnaire. When patients prescribed teratogenic medications were analyzed separately, there was statistically significant ($P < 0.0001$) better documentation of pregnancy planning after the intervention. Secondary analyses found that patient age, race and ethnicity, encounter type, and provider sex had no significant impact on documentation rates.

Conclusion. By integrating an electronic, previsit questionnaire into the patient portal, documentation was significantly improved for contraception and pregnancy planning. The results were sustained for six months. Further studies are needed to see if improved documentation translates into more effective reproductive health care discussions, referrals to gynecology, and subsequent improvement in reproductive health outcomes.

INTRODUCTION

Autoimmune and inflammatory rheumatic diseases (AIRDs) such as systemic lupus erythematosus and rheumatoid arthritis primarily affect female patients during their reproductive years. Poorly controlled disease at the time of conception, regardless of diagnosis, is associated with higher rates of adverse pregnancy outcomes. Data support improved pregnancy outcomes when disease activity is low or in remission and when pregnancies are planned.[1–3] Moreover, many of the medications used to control rheumatologic disease are teratogenic and contraindicated during pregnancy. Reproductive health counseling at each patient's initial and subsequent visits is recommended by the American College of Rheumatology and imperative to ensure providers are aware of patients' reproductive health goals and contraceptive use.[1] However, the most effective means to initiate reproductive health care discussions and readdress them over time is unclear.

With recent legislative changes in the United States, options for unintended pregnancies may be limited depending on the state of residence. Almost 50% of pregnancies are unplanned, with a higher frequency occurring in underserved patient populations.[4] Therefore, discussions regarding reproductive health goals and effective contraception are of the utmost importance. Findings from a 2021 qualitative study by Wolgemuth et al[5] highlight that women with rheumatic diseases want their provider, specifically their rheumatologist, to initiate conversations

[1]Benjamin Moon, MD, PhD: University of Texas Southwestern Medical Center, Dallas, and University of California, San Francisco; [2]Puneet Bajaj, MD, MPH, James Roberts, MD, Kathryn Dao, MD, Bonnie L. Bermas, MD, Brooke S. Mills, MD: University of Texas Southwestern Medical Center, Dallas; [3]Megan Clowse, MD, MPH: Duke University School of Medicine, Durham, North Carolina.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/acr.25562).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25562.

Address correspondence via email to Brooke S. Mills, MD, at Brooke.Mills@utsouthwestern.edu.

Submitted for publication October 25, 2024; accepted in revised form April 17, 2025.

SIGNIFICANCE & INNOVATIONS

- Quality care gaps exist regarding documentation of contraception and pregnancy planning among rheumatology providers.
- This study found an automated previsit questionnaire integrated into the electronic health record can improve the quality care gap.
- Implementation of the electronic previsit questionnaire improved reproductive health care documentation for patients regardless of race and ethnicity, provider sex, or type of visit.

regarding sexual and reproductive health and revisit the conversation periodically.

Several barriers limit providers from addressing contraception and reproductive health care on a regular basis. Lack of ownership of this health care issue, lack of knowledge regarding family planning, provider discomfort, and/or patient resistance in discussing sensitive topics have been identified as some potential barriers. Additionally, multiple studies have identified insufficient encounter time as the leading barrier to reproductive health care discussions.[6,7] The goal of this quality improvement (QI) project is to improve documentation of patients' pregnancy intent and contraceptive use among rheumatology providers at a single academic medical center from the baseline of 15.5% and 44.6%, respectively, to more than 50% and 70%, respectively, in six months.

MATERIALS AND METHODS

This project was completed at the University of Texas Southwestern Medical Center with data obtained from four different rheumatology clinics, including 17 health care providers. Institutional review board exemption was obtained for this QI initiative that involved minimal risk to patients but sought to improve health care outcomes. Charts were randomly selected and reviewed from female patients aged 18 to 45 years with an AIRD according to documentation. Preintervention data were collected as part of a multisite Rheumatology Research Foundation–funded project to improve contraception documentation. Data were obtained via retrospective chart review of 30 randomly selected charts per month from May 2022, August 2022, September 2022, January 2023, and March 2023. Selected charts were limited to follow-up visits for patients with established AIRD. A maximum number of three charts per provider per month were included to avoid oversampling of individual providers. Documentation was defined as a description of contraception use anywhere within the note, including visits in which the patient declined to answer.

In June 2023, a reproductive health assessment was added to the previsit questionnaire required to be completed by patients before each visit. The reproductive Health Assessment Questionnaire (HAQ) included three questions: (1) Are you sexually active? (2) Would you or your partner like to become pregnant in the next year? and (3) Which birth control are you using? Patients were able to decline answering any or all questions. Other items on the previsit questionnaire include the HAQ and the Clinical Disease Activity Index (CDAI) questions. Patients were sent an email reminder to fill out the questionnaire on the patient portal one week before their visit. If it was not completed, patients were given a paper copy of the questions for the medical assistant to add to the portal.

The previsit questionnaire responses were viewable by the rheumatology provider under a specific tab if they were using the recommended note template in the electronic health record. Typical workflow involved clicking on a "Rheumatology Assessments" tab to review the HAQ, CDAI, and reproductive health questions. Tender and swollen joint counts were added if applicable, and then this information was automatically pulled into the note. For "copied forward" notes, previous responses to the questions may be in the note if the tab is not clicked and refreshed every visit. Screenshots of the reproductive health questionnaire as viewed by the patients and the provider notes can be seen in Figure 1A and B, respectively.

Postintervention data were collected between July 2023 and December 2023 by selecting 30 random charts per month. Clinical data were collected by the authors and deidentified within a secure document for analysis.

Demographic information before and after implementation of the questionnaire were collected, as shown in Supplementary Table 1, including age, race, ethnicity, diagnosis, pregnancy status, pregnancy desire, sexually active ("yes" or "no"), type of contraception if applicable, current use of teratogenic medications, pregnancy planning if applicable, type of visit (telehealth vs in-person visit), and provider sex (male vs female). Medications considered to be teratogenic include methotrexate, mycophenolate, mycophenolic acid, cyclophosphamide, leflunomide, and thalidomide.

The primary outcome measures were obtained from documentation within providers' notes of contraception and pregnancy planning. The study was powered at 95% to detect an absolute increase in documentation rate of 20% or greater from baseline occurrence of 40% or less.

A chi-square test was performed for statistical analysis, with a P value of ≤ 0.05 being considered statistically significant unless otherwise stated. For analysis of contingency tables in which one or more cells contained fewer than 10 individual samples, a Fisher's exact test was performed instead. Figures were created, and statistical analyses were conducted using Graphpad Prism software.

RESULTS

Data from a total of 148 randomly selected charts before the intervention were collected as detailed in the Methods

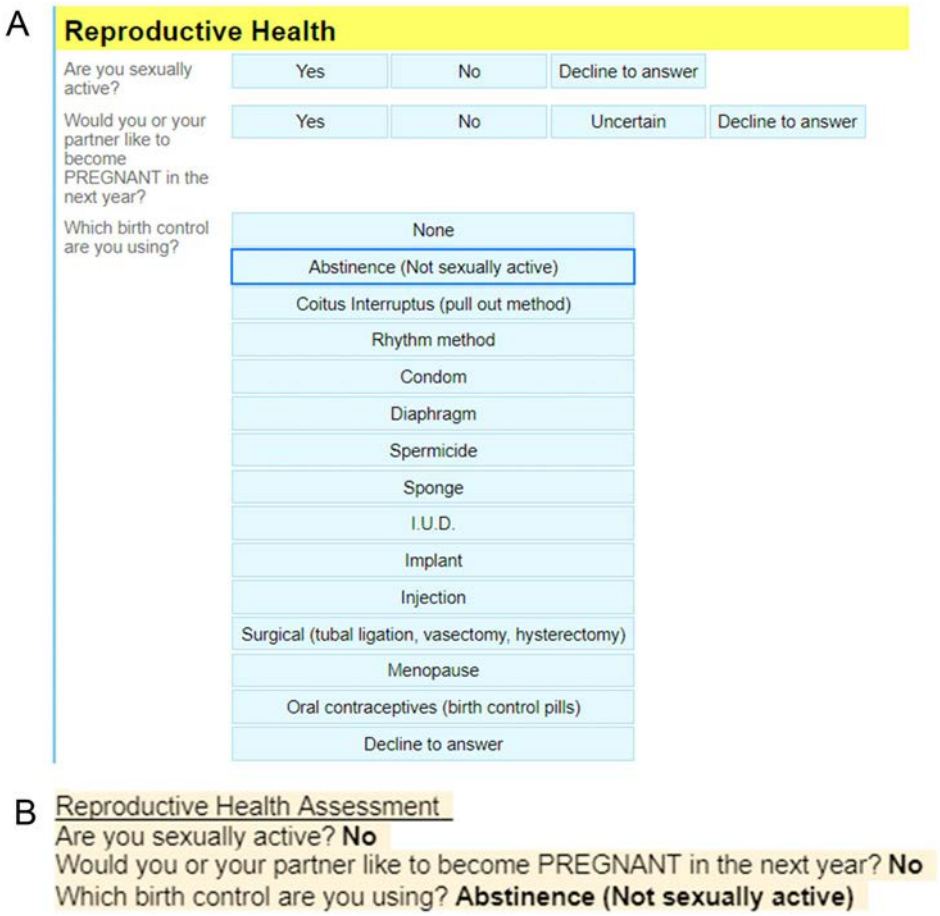


Figure 1. Electronic health questionnaire (2024 Epic Systems Corporation). (A) Screenshot of reproductive health questionnaire as seen by patients during previsit check-in. (B) Screenshot of autopopulated note template as seen by providers as seen in the electronic health record note.

section. Following the intervention, we sampled an additional 176 charts in the same manner. Two preintervention and four postintervention charts were excluded because they exceeded the predefined age cutoff. Demographic information was collected as outlined in the Methods section. There was no difference preintervention versus postintervention in race, ethnicity, encounter type, provider sex, or diagnosis. A significant difference ($P < 0.05$) was found between preintervention and postintervention age, teratogenic medication use, sexual activity, contraception type, and pregnancy desire. The difference in age was less than half an SD. The difference in teratogenic medication use, sexual activity, contraception type, and pregnancy desire were heavily influenced by the questionnaire and improved documentation of those variables after intervention (Supplementary Table 1).

Electronic questionnaire improving reproductive health documentation. Preintervention contraception use documentation averaged 44.6% of notes (Figure 2A and B). Postintervention documentation rates improved to an average of 70.5% of notes ($P < 0.0001$), with an immediate increase within one month that was sustained throughout the study duration. A divisional grand rounds presentation given in October of 2022 may account for the uptrend in contraception documentation rates seen before the initiation of the intervention.

Similarly, documentation of pregnancy intent was found in 15.5% of notes before implementation of the questionnaire. After intervention, documentation rates improved to 60.2% ($P < 0.0001$) (Figure 2C and D). The improvement was sustained through the postintervention period.

Electronic questionnaire detecting more at-risk patients. From the sampled charts, 51 preintervention and 43 postintervention patients were prescribed a predefined teratogenic medication (methotrexate, mycophenolate mofetil, or leflunomide). Among patients prescribed teratogenic medications, documentation of contraception use improved absolutely from 51% to 65.1% after intervention; however, this was not statistically significant ($P = 0.1673$) (Figure 3A). Documentation of pregnancy desire in this population also improved from 15.7% to 55.8% ($P < 0.0001$) (Figure 3B). Baseline documentation rates for pregnancy desire before intervention were not notably higher in patients taking teratogenic medications versus those who were not.

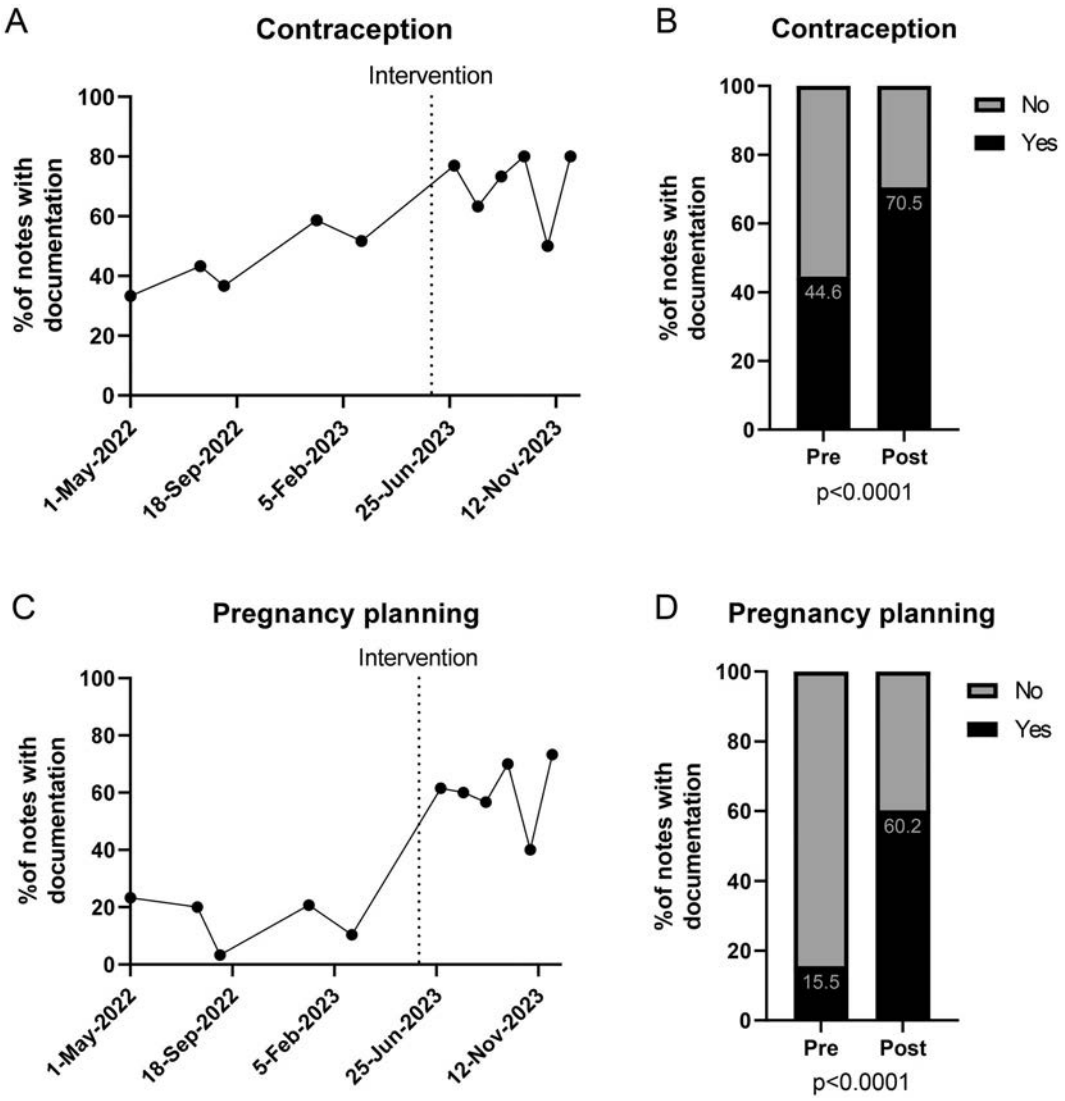


Figure 2. Contraception and pregnancy planning documentation rates before and after intervention. (A) Longitudinal documentation rates of contraception within providers' notes over the study period. Each data point represents one month of chart review. Questionnaire implementation marked by dotted line. (B) Aggregate contraception documentation rates before and after intervention. (C) Longitudinal documentation rates of pregnancy planning within providers' notes over the study period. Each data point represents one month of chart review. Questionnaire implementation marked by dotted line. (D) Aggregate pregnancy planning documentation rates before and after intervention.

Notes of high-risk patients who were taking teratogenic medications and who were sexually active were further analyzed. Before intervention, 26.9% of 26 high-risk patient notes had documentation of both contraception and pregnancy intention. After intervention, documentation improved to 62.5% of 24 high-risk patient notes. We identified a total of eight patients who were sexually active and not on contraception after intervention compared to one patient before intervention (P = 0.0103). Additionally, we identified two patients desiring pregnancy after intervention compared to one patient before intervention. Although the absolute number of patients found with these characteristics was low, these data indicate that the intervention helped identify more high-risk patients requiring discussions of care.

Age, race and ethnicity, and provider sex not affecting documentation rates. Additional secondary analyses were conducted to see if other factors influenced provider documentation. Within the postintervention sample, there was no significant difference in documentation of pregnancy planning when stratified based on patient age (≤30 years vs >30 years), race and ethnicity (White and not Hispanic vs Hispanic or another racial minority), or health care provider sex (male vs female) (Supplementary Table 2). There was an absolute but not statistically significant difference of 13.4% between telehealth and face-to-face encounters, suggesting telehealth encounters may have poorer documentation or limitations in time or connectivity (P = 0.1207).

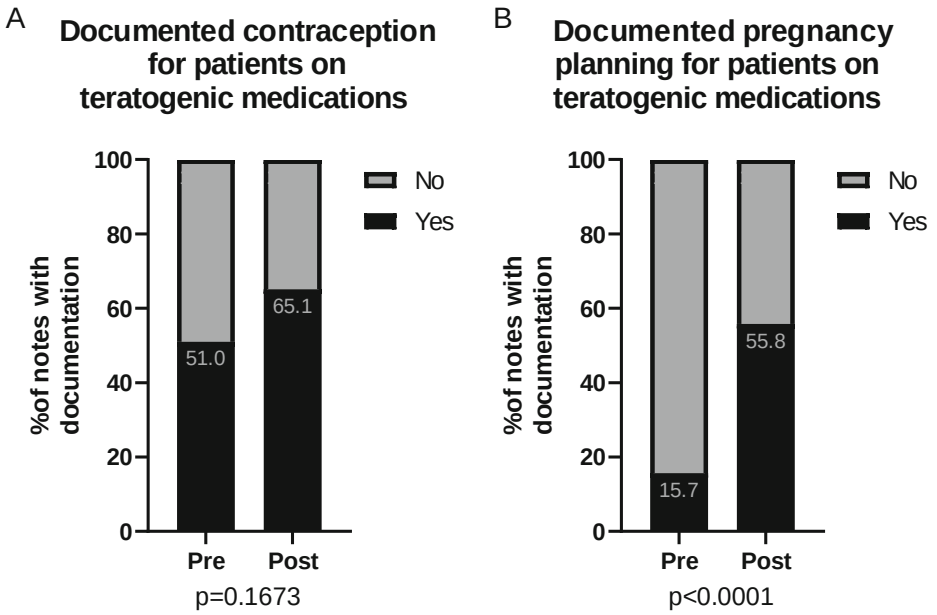


Figure 3. Reproductive health documentation of patients on teratogenic medications before and after intervention. (A) Contraception documentation rates between patients taking teratogenic medications (methotrexate, mycophenolate mofetil, mycophenolic acid, and leflunomide) before and after intervention. (B) Pregnancy planning documentation rates between patients taking teratogenic medications before and after intervention.

DISCUSSION

In this QI project, baseline data from rheumatology encounters showed a contraception documentation rate of less than 50% and a pregnancy intent documentation rate of 15%. After initiation of the reproductive HAQ, statistically significant improvement in contraception and pregnancy intent documentation was seen with rates more than 70% and 60%, respectively. Among patients taking teratogenic medications, documentation of contraception use and pregnancy intent improved from 51% to 65.1% ($P = 0.1673$) and 15.7% to 55.8% ($P < 0.0001$) after intervention, respectively. This study's postintervention data did not demonstrate a difference between documentation rates of pregnancy planning and contraception in White patients who were not Hispanic versus Hispanic patients or patients of another racial minority and it did not show a difference based on patient age. There was no difference between the contraception and pregnancy intention documentation rate between male and female providers.

Other studies have corroborated our findings of poor contraception and/or reproductive health counseling in high-risk patient populations.[6,8–10] This study's preintervention contraception documentation rate of 44.6% is higher than documentation rates cited in previous studies.[9–11] The previously mentioned average baseline documentation rates found in this study could be attributed to contraception documentation "smart" phrases in preexisting note templates and a recent grand rounds in the fall of 2022 regarding contraception and pregnancy planning in patients with AIRD.

Few studies have investigated documentation rates of pregnancy intent among patients with AIRD. The low baseline documentation rate in this study is similar to rates identified in another recent study by Pryor et al.[7] Although Pryor et al[7] found an almost two-fold better documentation rate among female providers, this study showed no significant documentation rate difference based on provider sex.

A strength of this study is the development of an automated previsit questionnaire that is not provider driven. This study's simple previsit questionnaire is completed by the patient and/or the medical assistant rather than the provider. This study suggests if similar automated, patient-driven previsit questionnaires are broadly implemented, existing barriers to reproductive health care discussions, specifically time concerns, may be mitigated.

Previsit assessment tools may allow patients to feel more involved in their care and can improve the physician–patient relationship.[12] With this intervention, documentation of contraception and pregnancy intent does not mean reproductive health care was discussed during the visit. However, using an automated previsit questionnaire before visits may remind patients to inquire about and feel more comfortable expressing their reproductive health care concerns and desires. Clear documentation of patients' reproductive health goals also may prompt providers to initiate these conversations. However, it remains unclear whether the improvement in documentation of contraception and pregnancy planning translates into more effective reproductive health care discussions in clinics.

A recent retrospective study of 43 pregnant women with lupus identified 40% of those pregnancies as unplanned,

with unplanned pregnancies being more common in lower age, income, and education populations.[13] Another study found women of Hispanic ethnicity and Black and Asian race were less likely to have contraception documentation than White women.[11] Automated, patient-driven previsit questionnaires may mitigate implicit bias when discussions are initiated by providers or staff.[14] However, more studies are needed to identify the optimal means to overcome these health care inequalities.

This study has several limitations. Results are from a single academic center and may not be generalizable to other centers or private practice. Additionally, not all patients complete their questionnaire due to difficulties with online portals and may have limited literacy when answering the questions on paper. Hence, previsit questionnaires may be more effective for those patients who are more literate or technologically savvy and biased against those patients with limited access or understanding. Documentation of contraception and reproductive health care in the current study also relied on providers using the template that automatically imported the patient responses. Some providers use different templates, so patient responses, although still viewable, were not automatically imported to the note. If notes are copied forward and providers are not clicking on the rheumatology assessments tab, the previous patient responses rather than the most current responses will be in the note. Lastly, reproductive health care discussions conducted during the visit but not documented in the notes were not captured by these data. Conversely, documentation in notes does not mean that the provider and patient discussed reproductive health care during the visit.

Contraception and pregnancy planning discussions early and often for patients with AIRDs are recommended. However, the current documentation rate of reproductive health discussions is lower than expected and demonstrates a significant quality-of-care gap. This study's data suggest implementing an automated previsit questionnaire before visits may improve documentation of contraception and pregnancy planning. More studies are needed to determine whether improved documentation translates into more effective reproductive health care discussions and improved pregnancy outcomes.

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 Mills 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. Sammaritano LR, Bermas BL, Chakravarty EE, et al. 2020 American College of Rheumatology guideline for the Management of Reproductive Health in Rheumatic and Musculoskeletal Diseases. Arthritis Care Res (Hoboken) 2020;72(4):461–488.
2. Phansenee S, Sekararithi R, Jatavan P, et al. Pregnancy outcomes among women with systemic lupus erythematosus: a retrospective cohort study from Thailand. Lupus 2018;27(1):158–164.
3. Skorpen CG, Lydersen S, Gilboe IM, et al. Influence of disease activity and medications on offspring birth weight, pre-eclampsia and preterm birth in systemic lupus erythematosus: a population-based study. Ann Rheum Dis 2018;77(2):264–269.
4. Finer LB, Zolna MR. Declines in unintended pregnancy in the United States, 2008-2011. N Engl J Med 2016;374(9):843–852.
5. Wolgemuth T, Stransky OM, Chodoff A, et al. Exploring the preferences of women regarding sexual and reproductive health care in the context of rheumatology: a qualitative study. Arthritis Care Res (Hoboken) 2021;73(8):1194–1200.
6. Hill W, Holy R, Traiger G. EXPRESS: intimacy, contraception, and pregnancy prevention in patients with pulmonary arterial hypertension: are we counseling our patients? Pulm Circ 2018;10(4): 2045894018785259.
7. Pryor KP, Albert B, Desai S, et al. Pregnancy intention screening in patients with systemic rheumatic diseases: pilot testing a standardized assessment tool. ACR Open Rheumatol 2022;4(8):682–688.
8. Joerns EK, Mills B, Makris UE, et al. Low rates of reproductive counseling documentation in women with interstitial pneumonia with autoimmune features. J Clin Rheumatol 2023;29(3):145–150.
9. Yazdany J, Trupin L, Kaiser R, et al. Contraceptive counseling and use among women with systemic lupus erythematosus: a gap in health care quality? Arthritis Care Res (Hoboken) 2011;63(3):358–365.
10. Birru Talabi M, Clowse MEB, Blalock SJ, et al. Contraception use among reproductive-age women with rheumatic diseases. Arthritis Care Res (Hoboken) 2019;71(8):1132–1140.
11. Clowse MEB, Li J, Talabi MB, et al. Frequency of contraception documentation in women with systemic lupus erythematosus and rheumatoid arthritis within the Rheumatology Informatics System for Effectiveness Registry. Arthritis Care Res (Hoboken) 2023;75(3): 590–596.
12. Gholamzadeh M, Abtahi H, Ghazisaeeidi M. Applied techniques for putting pre-visit planning in clinical practice to empower patient-centered care in the pandemic era: a systematic review and framework suggestion. BMC Health Serv Res 2021;21(1):458.
13. Rajendran A, Eudy AM, Balevic SJ, et al. The importance of pregnancy planning in lupus pregnancies. Lupus 2021;30(5):741–751.
14. Sims CA, Eudy AM, Doss J, et al. The impact of pregnancy planning and medical readiness on reproductive outcomes in women with systemic lupus erythematosus. Lupus 2023;32(14):1666–1674.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1246–1253
DOI 10.1002/acr.25552

BRIEF REPORT

Development, Usability, and Validity Evidence of a Rheumatology Telehealth Feedback Form

Lisa Zickuhr,[1] Alberto Sobrero,[1] Daniel Albert,[2] Amanda S. Alexander,[3] Tami Bonnett-Ami,[4] Sarah Dill,[5] Sharon Dowell,[6] Elizabeth D. Ferucci,[7] Connie Herndon,[8] Bharat Kumar,[9] David Leverenz,[10] Jennifer Mandal,[11] Amaad Rana,[12] Irene J. Tan,[13] Swamy Venuturupalli,[14] Tiffany Westrich-Robertson,[15] Marcy B. Bolster,[16] and Jason Kolfenbach,[17] for the American College of Rheumatology Workforce Solutions Committee

Objective. Rheumatology telehealth is widespread, making it essential that rheumatology fellows-in-training (FITs) achieve competence delivering telehealth care before entering the workforce. Feedback enhances telehealth skill development. This study develops a Rheumatology Telehealth Feedback Form (RTFF) that incorporates existing data and expertise as well as gathers validity evidence supporting its use.

Methods. The American College of Rheumatology Workforce Solutions Committee formed a working group with expertise in rheumatic diseases, telehealth, and medical education. The working group conducted a modified Delphi and focus groups to define content grounded in Rheumatology Telehealth Competencies and evidence-based practice as well as develop the RTFF. Rheumatology educators and FITs piloted the RTFF during patient care and simulated telehealth encounters, rating the helpfulness of the RTFF and providing data to calculate intraclass correlation coefficients (ICCs).

Results. The modified Delphi identified 16 skills essential to conducting rheumatology telehealth encounters, which were converted into 17 items on the RTFF. Comment boxes prompted supplementary narrative. Five educators and 10 FITs piloted the materials. All educators (5 of 5) "agreed" or "strongly agreed" that the RTFF helped teach rheumatology telehealth skills. ICCs calculated from two simulated telehealth encounters demonstrated moderately to strongly supportive interrater reliability (ICC = 0.857 and 0.632, respectively).

Conclusion. The RTFF incorporates an evidence-based, expert consensus of the skills most essential for rheumatology telehealth encounters. Educators can use the RTFF as a guide for observing FITs delivering telehealth care, providing formative feedback, and helping FITs develop their skills before entering the workforce. The RTFF is the first direct observation formative assessment tool designed for teaching rheumatology telehealth skills.

Supported by a Pfizer, Inc education grant dedicated to enhancing telehealth practice among providers and experience among people with rheumatic diseases (76770947).

[1]Lisa Zickuhr, MD, MHPE, Alberto Sobrero, BS: Washington University in St. Louis School of Medicine, St. Louis, Missouri; [2]Daniel Albert, MD: Dartmouth College Geisel School of Medicine, Lebanon, New Hampshire; [3]Amanda S. Alexander, MD: University of Alabama at Birmingham Marnix E. Heersink School of Medicine; [4]Tami Bonnett-Ami, BA: 3GEM Consulting, Washington, DC; [5]Sarah Dill, MD: Lakeview Rheumatology, Yakima, Washington; [6]Sharon Dowell, MBBS, MSc: NEA Baptist Memorial Hospital, Jonesboro, Arkansas; [7]Elizabeth D. Ferucci, MD, MPH: Alaska Native Tribal Health Consortium, Anchorage; [8]Connie Herndon, BA: American College of Rheumatology, Atlanta, Georgia; [9]Bharat Kumar, MD, MME: University of Iowa Carver College of Medicine, Iowa City; [10]David Leverenz, MD, MEd: Duke University School of Medicine, Durham, North Carolina; [11]Jennifer Mandal, MD: University of California San Francisco School of Medicine; [12]Amaad Rana, MD: Yale School of Medicine, New Haven, Connecticut; [13]Irene J. Tan, MD: Einstein Philadelphia Jefferson Health, Philadelphia, Pennsylvania; [14]Swamy Venuturupalli, MD: Cedars Sinai Medical Center, Beverly Hills, California; [15]Tiffany Westrich-Robertson, BA: International Foundation for Autoimmune & Autoinflammatory Arthritis, St. Louis, Missouri; [16]Marcy B. Bolster, MD: Massachusetts General Hospital, Boston; [17]Jason Kolfenbach, MD: University of Colorado Anschutz School of Medicine, Aurora.

Drs Bolster and Kolfenbach are co-last authors and contributed equally to this work.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr.25552).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25552.

Address correspondence via email to Lisa Zickuhr, MD, MHPE, at zickuhrl@wustl.edu.

Submitted for publication September 19, 2024; accepted in revised form April 8, 2025.

SIGNIFICANCE & INNOVATIONS

- Rheumatology telehealth represents an important form of care delivery to people with rheumatic diseases.
- As future entrants into the rheumatology workforce, fellows-in-training (FITs) must develop competence in the skills for delivering telehealth care, a process that benefits from observation and formative feedback at the point of care.
- Rheumatology health care providers, educators, and patients achieved consensus on the skills most essential to telehealth encounters and incorporated the items into a Rheumatology Telehealth Feedback Form (RTFF).
- The RTFF supports FITs' development of rheumatology telehealth skills by guiding educators to observe FITs as they conduct telehealth encounters and provide formative feedback.

INTRODUCTION

Rheumatology telehealth (rTH), or the use of telecommunications for delivering care to people with rheumatic diseases, enhances clinical practice.[1] rTH expands opportunities for accessing care and addresses rheumatology workforce shortages, especially for people limited by travel.[2] Furthermore, rTH supports high-quality care and improves health care decisions, access, and cost savings while maintaining patient satisfaction.[3,4] It is critical that rheumatology training ensures graduating fellows-in-training (FITs) competently deliver high-quality rTH care as new entrants into the workforce.

Direct observation (DO) supports the development of FITs' skills. In DO, supervising faculty watch FITs deliver patient care and provide feedback that advances learning.[5] Tools, such as the mini-Clinical Evaluation Exercise, focus educators' observations on the skills FITs implement during in-person patient care.[5] These tools also prompt educators to deliver feedback that promotes development.

At present, no peer-reviewed tools exist that guide educators in providing DO and feedback to FITs during rTH. Previously published resources include the Rheumatology Telehealth Competencies, classroom-based education materials, and patient simulations.[6,7] Comprising 24 skills unique to rTH, the Rheumatology Telehealth Competencies reflect the knowledge, attitudes, and skills contributing to telehealth care delivery.[6] The Rheumatology Telehealth Competencies are a useful resource when considering the large scope of rTH instruction, but they contain more than educators can assess when teaching at the "webside."

In an independent quality improvement initiative, a member of our team (MBB) surveyed educators' needs regarding rTH educational materials before our study. A total of 45 of 150 educators (30%) answered 10 questions surveying their rTH educational needs. A total of 25 of 43 educators (56%) reported needing additional rTH curricular elements, with 18 of 43 respondents (42%) requesting additional rTH assessment tools (unpublished data).

This study aims to develop a Rheumatology Telehealth Feedback Form (RTFF) that facilitates DO and delivery of feedback to FITs conducting telehealth encounters. Our methods primarily seek expert consensus regarding content that grounds the RTFF in established Rheumatology Telehealth Competencies and evidence-based practice. Additionally, pilot implementation of the RTFF reflects the form's usability and validity evidence.

MATERIALS AND METHODS

The American College of Rheumatology Workforce Solutions Committee invited rheumatology providers, educators, and a patient with expertise in telehealth to join the RTFF Working Group. Participants included 14 individuals representing all four geographic regions of the United States (Northeast, South, Midwest, and West) as well as academic and community practice. Working group members provided waived consent for completing surveys and verbal consent for participating in video teleconference meetings in accordance with the Declaration of Helsinki, as approved by the Washington University in St. Louis School of Medicine Institutional Review Board (no. 202304097 and 202310110).

Expert consensus on content. We implemented modified Delphi methods to identify content for the RTFF, setting the threshold for consensus at 70% and incorporating inductive reasoning during a video teleconference meeting. A subset of the working group (LZ, MBB, and JK) referenced the Rheumatology Telehealth Competencies, literature, and their experience to list skills integral to rTH encounters within the domains of webside manner, physical examination and objective findings, shared decision-making, and visit coordination.[2–6] The working group prioritized defining the skills most essential and specific to rTH encounters rather than comprehensively addressing all skills used during patient care. Therefore, items focused on skills that providers implement uniquely or with nuance during rTH care; tasks routinely completed in person or by office staff were excluded.

Via online surveys (Qualtrics), working group members (LZ, DA, ASA, S Dill, S Dowell, EDF, BK, DL, JM, IJT, SW, TWR, MBB, and JK) rated the essential nature of each item for conducting rTH encounters according to a Likert scale (not at all essential, somewhat essential, essential, and highly essential). Selections of "essential" and "highly essential" designated item inclusion. The survey also prompted participants to submit items they thought were missing from the initial list. Participants completed two rounds of Delphi surveys; the second round included new items submitted in the first round. The working group met via video teleconference to discuss items not yet meeting consensus and to

review the skills deemed essential to rTH encounters, at times editing items.

Feedback form development. The content generated from the modified Delphi guided the development of the RTFF and an instruction guide. Working group members with specific expertise in rheumatology education (LZ, ASA, BK, DL, IJT, MBB, and JK) reviewed the RTFF and instruction guide via video teleconference, refining the materials.

Usability data. Five rheumatology educators external to the working group piloted the RTFF and instruction guide after viewing recordings of mediocre and exemplary telehealth encounters.[6] These faculty completed RTFFs for the recorded encounters while providing feedback on the RTFF form and instruction guide via an online survey. Two working group members (LZ and AS) conducted qualitative content analysis of educators' narrative comments to identify themes from this feedback,[8] which guided improvement of the RTFF and instruction guide. We calculated intraclass correlation coefficients (ICCs) to measure interrater reliability using RTFF total scores from the recorded encounters (SPSS, IBM).

Next, faculty educators from four academic medical centers observed five first- and five second-year FITs conduct video teleconference and audio-only encounters in the ambulatory setting, scored the FITs' skills against the RTFF, and used the RTFF to deliver immediate feedback (LZ, ASA, DL, AR, and JK). We used an independent two-way *t*-test to compare the number of "not applicable" selections between the video teleconference and audio-only DOs as an estimate of the RTFF's usefulness in both rTH formats. The FITs completed an online survey describing the clarity and helpfulness of the RTFF. FITs provided verbal consent for their participation in the research study, and patients permitted observations during their rTH appointments.

RESULTS

Modified Delphi. The working group identified 34 items as essential components of rTH encounters through Delphi surveys. During the video teleconference meeting, the group reduced content to 16 actions (Supplement 1).

Members adjusted items to increase their specificity for rTH. They eliminated some statements that met the Delphi's threshold for inclusion because the skill also applied to in-person encounters, such as "Involves care partners in decision-making." Items with a nuanced element for telehealth, for instance, the skill "Establishes or reaffirms rapport," were included because the process for these items have different features during telehealth visits compared to in-person encounters.[9] The working group similarly revised the item referencing patient reported outcomes to better reflect telehealth workflows and published recommendations for rTH care.[10]

Members prioritized content reflecting common telehealth practices. Within the domain of physical examination and objective data, the group excluded skills pertaining to integrating telehealth presenters because this resource is uncommonly used.

Lastly, members consolidated items to avoid redundancy and enhance ease of observation. Multiple items within webside manner and shared decision-making were grouped under headings collectively describing the complex skills of verbal communication and sharing decisions during telehealth.

RTFF and instruction guide. The design of the RTFF (Figure 1) and instruction guide (Supplement 2) incorporates best practices for delivering feedback, including referencing the specifics of the patient encounter, reinforcing skills performed well, and identifying opportunities for improvement.[11] A three-point scale, ranging from "not competent" to "fully competent" simplifies observers' scoring while also highlighting skills that were done well or need improvement. The working group selected the adjective "competent" for the scale because they thought this description would be grasped easily. A "not applicable" descriptor allows educators to disregard items irrelevant to the observed encounter. The instruction guide supplements the RTFF, providing definitions of the anchor headings and instructions for scoring.

The RTFF categorizes content from the modified Delphi into four sections: (1) webside manner, (2) physical examination and objective findings, (3) shared decision-making, and (4) visit coordination. For items representing complex behaviors or skills with nuance during telehealth, parenthetical examples of ways FITs may demonstrate the complex skill reduce cognitive load, preventing the observer from having to define these elements themselves. The instruction guide provides additional examples of how each skill might be implemented during rTH.

Boxes prompting narrative comments conclude each section. Herein, educators expand on what was done well while describing areas for improvement with recommendations for FITs to incorporate feedback. These narratives provide description and actionable comments directing growth. The RTFF prints on one page (front and back) to optimize DO of FITs rather than diverting attention to a multipage form.

Themes emerged from the likes and dislikes the pilot educators submitted regarding the RTFF and instruction guide, including describing the assessment data collected (the RTFF's subcategories and Likert scale), content (items' wording and RTFF length), instruction (the instruction guide and examples), and the RTFF's clinical relevance (Table 1). A subset of the working group (LZ, MBB, and JK) applied these comments to improve the RTFF and instruction guide, making modifications that enhanced the clarity of items, divided communication skills into two items describing active listening and verbal telehealth communication skills, and added an example about the virtual

Date:	Learner Name:	Observer Name:
Instructions to the Observer: Observe the learner conduct a telehealth encounter and provide feedback using this form, completing the checklist items and writing additional comments in the designated areas.		

	Not Competent (0 points)	Somewhat Competent (1 point)	Fully Competent (2 points)	Not Applicable
"Webside" Manner				
Sets up the workstation to enhance communication (e.g., camera and microphone positioning)				
Orients patient +/- care partner to the telehealth platform (e.g., videoconference platform features, contingency plan for lost connection) if not already done by clinic staff				
Establishes or reaffirms rapport with patient +/- care partner				
Determines and adjusts when the current encounter is unsuitable/unsafe for telehealth (e.g., converts to emergent or urgent in-person visit)				
Assesses patient's environment (e.g., for safety and privacy as well as to glean information pertinent to clinical care)				
Implements telehealth active listening skills (e.g., maintains eye contact, interprets and uses nonverbal cues, balances talking with listening, and summarizes as is fitting for the telehealth scenario)				
Implements telehealth communication skills (e.g., conveys clear messages, verifies patient understanding, and utilizes interpreter services as is fitting for the telehealth scenario)				
Additional Comments on "Webside" Manner:				

	Not Competent (0 points)	Somewhat Competent (1 point)	Fully Competent (2 points)	Not Applicable
Physical Exam and Objective Findings				
Informs patients when the telehealth physical exam may be insufficient to evaluate symptoms				
Modifies relevant physical exam maneuvers to the telehealth platform, including soliciting patient-generated images when photographs better represent findings				
Provides clear instructions to patient +/- care partner for conducting the virtual physical exam				
Incorporates patient-reported indices of disease activity as part of longitudinal disease management				
Additional Comments on Physical Exam and Objective Findings:				

Figure 1. The rheumatology telehealth feedback form.

	Not Competent (0 points)	Somewhat Competent (1 point)	Fully Competent (2 points)	Not Applicable
Shared Decision-Making				
Identifies scenarios when decisions are better made in-person than virtually (or vice versa)				
Utilizes features of the telehealth platform (e.g., screen sharing) when they may facilitate shared decision-making				
Adapts shared decision-making to telehealth (might include admitting clinical uncertainty from limitations of telehealth or extending discussion across multiple encounters to provide patients adequate time for decision-making)				
Additional Comments on Shared Decision-Making:				

	Not Competent (0 points)	Somewhat Competent (1 point)	Fully Competent (2 points)	Not Applicable
Visit Coordination				
Creates a telehealth care plan that navigates potential social and structural barriers to care				
Coordinates after-visit tasks for the patient with members of the health care team (e.g., schedulers, social workers, nurses)				
Determines the location and timing of the next encounter, congruent with the level of patient's disease acuity (e.g., in-person vs virtual follow-up)				
Additional Comments on Visit Coordination:				

Encounter Details		
Primary Encounter Diagnosis:		
Was the observed encounter simulated?	□ Yes	□ No
What type of encounter was observed? Mark all that apply.	□ Initial visit □ Subsequent visit	□ Audio only □ Video teleconference
What was discussed during the encounter? Mark all that apply.	□ Acute symptom □ Routine follow-up	□ Test results □ Diagnosis education □ Treatment planning

Figure 1. (Continued)

Table 1. Thematic analysis with illustrative quotes from educators' comments after piloting the rheumatology telehealth feedback form and instruction guide

Themes	Likes and dislikes	Illustrative quotes
Assessment data: the form's subcategories, data metrics	Likes: • Intuitiveness of graded responses • Space for narrative feedback. Dislikes: • Excessive denseness of data associated with the webside manner section.	(+) "The scoring scale was fairly intuitive" (+) "I like the space included for comments within each section" (−) "I think the communication skills category…is a little dense for one item…."
Content-specific items, form length	Likes: • Comprehensive prompts • Appropriate length Dislikes: • Vague prompts • Prompts too comprehensive	(+) "…the items are comprehensive in assessing the important features of a successful telemedicine encounter." (−) "Determines whether the current encounter is suitable for telehealth assessment' was a little vague in terms of what we were actually supposed to be scoring." (−) "The item about coordination with health care team was a little confusing…. I assume you meant nursing or front desk."
Instruction: examples, instruction guide	Likes: • Instructions document • Helpful examples Dislikes: • Not enough example variety	(+) "The instructions document was very detailed and provided strong examples that I could use as a reference when grading the videos." (−) "The examples given for 'identifies scenarios when decisions are better made in-person' should include something about inability to…perform a comprehensive physical exam."
Clinical relevance: relevance to clinical care	Likes: • Content relevant to clinical practice. • Helpful for feedback delivery.	(+) "I think the skills assessed were very important and wish I had training on how to conduct telehealth visits well during my fellowship."

physical examination limiting shared decision-making to the instruction guide.

Usability. In the pilot implementation, five educators applied the RTFF against prerecorded excellent and mediocre rheumatology telehealth encounters and achieved ICCs supportive of high and moderate reliability (ICC = 0.857 and 0.632, respectively). Additionally, they shared favorable opinions regarding the RTFF and the instruction guide. All educators (5 of 5) "agreed" or "agreed a lot" that the instruction guide is clear and the examples are helpful. Similarly, all educators "agreed" or "agreed a lot" that the RTFF is easy to use, directs them to observe important skills, and provides meaningful feedback to FITs. One educator commented on the combined novel and essential nature of rTH education, stating, "I think the skills assessed were very important and wish I had training on how to conduct telehealth visits well during my fellowship." Prompts related to the clarity of items on the RTFF received the lowest average scores (3.4/4.0), resulting in the modifications discussed earlier. During DO of FITs, 3 of the 10 encounters featured audio-only visits with no difference in the number of items marked "not applicable" compared to video teleconference DOs (audio-only mean 4.33, video teleconference mean 4.25, P = 0.963), suggesting the RTFF's usability in both formats. Nine of the 10 FITs completed the survey (90%), all of whom (9 of 9) "agreed" or "strongly agreed" that the items of the RTFF are relevant to their practice, helped them identify areas for improvement, and provided high-quality feedback.

DISCUSSION

Rheumatology telehealth is an important form of care delivery to people with rheumatic diseases, making rTH skill development requisite for FITs. A full complement of educational materials that includes didactic, simulation, and assessment strategies best equip educators to teach rTH skills. In this study, we created the RTFF and instruction guide to meet the need for tools that measure and deliver feedback on FITs' skills in a way that is useful to both educators and trainees.

The incorporation of rTH into fellowship training is relatively new, having been introduced broadly during the COVID-19 pandemic.[9] Many faculty developed their skills independently during clinical practice and may not have a framework for teaching the knowledge, attitudes, and skills requisite for rTH. The RTFF and its instructional guide address this gap and provide a framework for rTH instruction at the webside.

The RTFF provides formative feedback to FITs to improve skill development. In contrast, summative assessment determines whether FITs' collective skillset meets established proficiency standards, achieving readiness for independent practice.[5,12,13] Summative assessment tools require extensive validity evidence and setting thresholds for competence that were not conducted when developing the RTFF. Therefore, the

RTFF's value lies in guiding educators to provide formative feedback to FITs on their rTH skills at the webside rather than measuring summative achievement.

DO, well-designed formative assessment tools, and feedback are of high educational value to trainees. These methods provide insight into learners' patient care skills by incorporating multiple skill domains, aligning observations with established competencies, and taking place in clinical (versus simulated) environments.[3,12,13] The RTFF and its instruction guide fulfill these parameters, as their development is grounded in the established literature with implementation at the point of care. The RTFF additionally supports quality rTH care because its content aligns with established telehealth quality metrics,[14] including the incorporation of disease activity measures as health outcomes and key performance metrics, emphasis on webside manner and shared decision-making as components of patient experience, and recognition of the clinical appropriateness of rTH encounters to ensure patient safety in care delivery. The addition of the RTFF to educators' resources for rTH expands opportunities for assessment and instruction beyond simulation to the ambulatory telehealth environment.

The instruction guide supplements the RTFF and reinforces standardization of application. The guide defines the Likert scale anchors and includes examples for how FITs may implement each item on the RTFF during DO, thereby directing educators' focus toward particular actions while providing suggestions for ways FITs might improve their rTH skills. The instruction guide may even be shared with FITs as an educational resource that lists different methods for implementing rTH skills.

Rigorous methodology reinforces the quality of the RTFF. A team of providers, educators, and a patient with expertise in rTH completed a modified Delphi to achieve consensus on content for the RTFF. This process enhances content validity and the trustworthiness that the RTFF assesses essential skills in rTH.[13] Additionally, the instruction guide helps educators apply the RTFF consistently regardless of previous experience with rTH or DO, reinforcing response process and the likelihood that the RTFF is being used as it was designed.[12] In a pilot, utilization of the RTFF produced moderately to strongly favorable interrater reliability scores to further support the tool's quality. These methods and results, combined with educators' and FITs' opinions that the RTFF is helpful, indicate the RTFF is a valuable tool for delivering formative, actionable, and timely feedback on FITs' rTH skills with benefit to both educators and trainees.

The limitations of our study include the small number of educators and FITs who piloted the RTFF. Although educators and FITs represented multiple programs, geographic locations, and stages of fellowship training, a larger pilot may have identified additional recommendations for improving the RTFF or demonstrated different levels of interrater reliability. The pilot generated strong interrater reliability (ICC = 0.857) when the RTFF was applied to a recording of an excellent rTH encounter but moderate reliability (ICC = 0.68) for the mediocre rTH encounter. Interrater reliability is very sensitive to sample size, and a larger pilot could have produced improved values. Additionally, pilot educators prepared their mental model for using the RTFF from the instruction guide alone, whereas group discussion could have further enhanced reliability measures. The tool could be refined to improve reliability; however, the utilization of the RTFF for formative feedback—as the working group intended its use—frees the tool from these limitations that are necessary for summative assessment. Furthermore, the working group included one person diagnosed with rheumatic disease. This individual was selected because of their extensive involvement with patient advocacy and their ability to represent their personal lived experiences alongside narratives other patients have shared with them. Nonetheless, incorporation of additional patients may have increased the patient-centeredness of the RTFF.

As future entrants into the rheumatology workforce, FITs must develop rTH skills. The RTFF supports and guides this process, providing an instrument for educators to deliver specific, timely, formative feedback at the webside. In doing so, educators prepare FITs to deliver rTH care and ensure the workforce continues to offer flexible modes of care delivery to people with rheumatic disease.

ACKNOWLEDGMENTS

We thank the educators and rheumatology fellows-in-training who piloted the Rheumatology Telehealth Feedback Form.

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 Zickuhr 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

Pfizer had no role in the writing of the manuscript or the decision to submit the manuscript for publication. Publication of this article was not contingent upon approval by Pfizer.

REFERENCES

1. Medicare Telehealth Trends. Centers for Medicare & Medicaid Services. 2024. Accessed July 24, 2024. https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-service-type-reports/medicare-telehealth-trends
2. Zickuhr L, McCarthy C, Nolan-Thomas E, et al. Shared decision-making during virtual care regarding rheumatologic and chronic

conditions: qualitative study of benefits, pitfalls, and optimization. ACR Open Rheumatol 2024;6(1):32–42.

3. Haleem A, Javaid M, Singh RP, et al. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117.
4. Venuturupalli S, Peck A, Jinka Y, et al. Home-based telemedicine in rheumatology-a scoping review. ACR Open Rheumatol 2024;6(5): 312–320.
5. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE Guide No. 31. Med Teach 2007;29(9):855–871.
6. Zickuhr L, Albert DA, Herndon C, et al; American College of Rheumatology Committee on Training and Workforce. Addressing competency in rheumatology telehealth care delivery. Arthritis Care Res (Hoboken) 2023;75(6):1213–1219.
7. Bolster MB, Kolfenbach J, Poeschla A, et al. Incorporating telemedicine in rheumatology fellowship training programs: needs assessment, curricular intervention, and evaluation. Arthritis Care Res (Hoboken) 2023;75(12):2428–2434.
8. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277–1288.
9. Bolster MB, Chandra S, Demaerschalk BM, et al; Virtual Care and Medical Educator Group. Crossing the virtual chasm: practical considerations for rethinking curriculum, competency, and culture in the virtual care era. Acad Med 2022;97(6):839–846.
10. England BR, Barber CEH, Bergman M, et al. Adaptation of American College of Rheumatology rheumatoid arthritis disease activity and functional status measures for telehealth visits. Arthritis Care Res (Hoboken) 2021;73(12):1809–1814.
11. Ross S, Hamza D, Zulla R, et al. Development of and preliminary validity evidence for the EFeCT feedback scoring tool. J Grad Med Educ 2022;14(1):71–79.
12. Prakash J, Chatterjee K, Srivastava K, et al. Workplace based assessment: a review of available tools and their relevance. Ind Psychiatry J 2020;29(2):200–204.
13. Criscione-Schreiber L. Turning objective structured clinical examinations into reality. Rheum Dis Clin North Am 2020;46(1):21–35.
14. Chuo J, Macy ML, Lorch SA. Strategies for evaluating telehealth. Pediatrics 2020;146(5):e20201781.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1254–1262
DOI 10.1002/acr.25557

American College of Rheumatology
Empowering Rheumatology Professionals

Ageism in Rheumatology: The Health Care Professional's Perspective

Aaron P. Smith,[1] Pooja M. Achanta,[2] Jiha Lee,[3] Namrata Singh,[4] and Una E. Makris[5]

Objective. Ageism (age-based stereotypes, prejudice, or discrimination) is prevalent and linked to prolonged disability and reduced lifespan in older adults. Little is known about ageism within rheumatology. This study explores the health care professional's (HCP) perception of the care of older adults and how ageist attitudes or perspectives may impact rheumatologic care.

Methods. A REDCap survey related to the clinical care of older adults that included the validated Expectations Regarding Aging (ERA-12) instrument (higher scores associated with less age-related bias) was administered to a convenience sample of HCP caring for patients with rheumatic disease. We calculated correlations between ERA-12 scores and the responses to other survey questions.

Results. A total of 255 surveys were collected from January 2023 to December 2023. Respondents were predominantly female (63%), White (70%), physicians (75%), healthcare professionals practicing in academic (66%) or in urban (64%) settings, and most practices having >25% adults over the age of 65 years (88%). The median ERA-12 score was 36 of 48, indicating that respondents, on average, disagreed with the stereotypes regarding aging. Higher ERA-12 scores were associated with greater enjoyment of the care of older adults ($P < 0.001$) and awareness of the Geriatric 5Ms (mind, mobility, medications, multicomplexity, and matters most ($P < 0.001$), a framework for improving age-friendly care. Lower ERA-12 scores were associated with believing that older adults are more demanding of attention ($P < 0.001$) and shifting from disease-modifying therapy to symptom relief in older adults ($P < 0.001$).

Conclusion. Stereotypical beliefs regarding aging are associated with self-reported changes to patient counseling and medical decision-making, suggesting age-related biases may affect the care of older adults with rheumatic diseases.

INTRODUCTION

The US population is rapidly aging, with the population of people older than 65 years increasing in size by one-third over the past decade.[1] As the rheumatology patient population ages, degenerative changes and progression and/or severity of rheumatic diseases may exacerbate the burden of disease and reduce quality of life.[2–5] Particularly common among the aging population is multicomplexity, polypharmacy, geriatric syndromes (eg, falls, delirium or dementia), frailty, and fragmented social support systems, all of which contribute to higher disease burden, disability, and costs.[6,7] There is an urgent need to educate the rheumatology field about aging principles and to adapt clinical practices to provide better care for older adults with rheumatic diseases.[8]

Change in practice requires cultural change, and ageism—stereotypes, prejudice, and discrimination based on age—is invisible yet ingrained and pervasive in our culture. Ageism affects billions of individuals, with approximately 50% of people worldwide holding ageist attitudes toward older adults and one-third of Europeans of all ages reporting experiencing ageism.[9] Ageism has been identified in every country and on every continent with an increased prevalence of ageist attitudes in developing countries.[10] Ageism can be self-directed or internalized, interpersonal, and/or institutional. Self-directed ageism is especially harmful because longitudinal studies have shown that patients with increased stereotypical beliefs regarding aging have worse medical outcomes up to decades in the future compared to age- and

National Institutes of Health, NIA P30 AG022845. Health Services Research and Development, HX003350, IIR 20-256.

[1]Aaron P. Smith, MD, MS: University of North Carolina at Chapel Hill; [2]Pooja M. Achanta, MD: Beth Israel Deaconess Medical Center, Boston, Massachusetts; [3]Jiha Lee, MD, MHS: University of Michigan, Ann Arbor; [4]Namrata Singh, MD, MSCI: VA Puget Sound Health Care System and University of Washington, Seattle; [5]Una E. Makris, MD, MSc: VA North Texas Health Care System and University of Texas Southwestern Medical Center, Dallas.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25557.

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

Submitted for publication October 23, 2024; accepted in revised form April 8, 2025.

SIGNIFICANCE & INNOVATIONS

- Increased stereotypical beliefs regarding aging are associated with increased focus on the risks of medical intervention rather than the benefits when counseling older adults with rheumatic diseases.
- Increased stereotypical beliefs regarding aging are associated with reduced enjoyment of the care of older adults and a shift in focus from disease-modifying therapy to symptomatic relief when treating older adults.
- Knowledge of the Geriatric 5Ms (mind, mobility, medications, multicomplexity, and matters most), was correlated with reduced stereotypical beliefs regarding aging.

health-matched controls with less negative views toward the aging process.[11–14]

Multiple studies have examined the link between ageism and health care outcomes, with >95% showing a significant negative effect.[10] The negative effects of ageism on physical and mental health include reduced lifespan, prolonged disability, and accelerated cognitive decline among other negative outcomes.[9] Additionally, a study of costs and prevalence of the eight conditions that most frequently affect older people (eg, cardiovascular disease, chronic respiratory disease) estimated that approximately 15% of total health care dollars spent on these conditions can be attributed to ageism. This amounts to approximately 63 billion dollars per year just in the United States.[11] The effects of ageism are likely to become even more pronounced over time, as a recent meta-analysis showed that health care professionals (HCPs) are increasingly demonstrating ageism in recent years, whereas no such trend exists for other professions.[10,15]

Previous work in rheumatology has also suggested that age-related bias impacts the care that older adults receive, especially in the care of rheumatoid arthritis (RA). Survey-based studies have shown that rheumatologists are significantly more likely to prefer less aggressive treatment of RA in older adults than in younger adults[16] and that age is weighted equally with the presence of erosions in the decision to escalate care in RA treatment.[17] Consistent with these findings, studies of patients with RA have consistently demonstrated significantly reduced use of biologics and targeted synthetic disease-modifying antirheumatic drugs (DMARDs) in the older adult patient population, with similar differences seen in the initial treatments chosen for young-onset RA versus late-onset RA.[18–25]

Despite these previous findings, age-related bias remains an underappreciated driver of treatment patterns in the care of patients with rheumatic disease. This is especially significant because undertreatment of rheumatic disease in this vulnerable population would be expected to lead to increased functional limitations and contribute to age-related disability, which is highly relevant to the individual, their caregivers, and society. The objective of this survey-based study was to gain a better understanding of the rheumatology HCP's perception of the care of older adults and how ageist attitudes or perspectives may impact rheumatologic care.

PATIENTS AND METHODS

All HCPs providing care for patients with rheumatic diseases were eligible to complete the survey. HCPs including board-certified rheumatologists, rheumatology fellows, advanced practice providers, physical therapists, nurses, pharmacists, social workers, medical students, psychologists, and chiropractors were eligible to complete the survey. The investigators invited and sought to gather an interprofessional and interdisciplinary perspective, with the requirement that respondents must care for patients with rheumatic disease. HCPs who do not care for patients with rheumatologic conditions, do not speak English, or who do not have internet access were not included in this study. This study was approved by the UT Southwestern Medical Center Institutional Review Board (STU-2022-0852).

We created a REDCap English-language survey containing 10 demographic questions, 12 questions from the Expectations Regarding Aging (ERA-12) instrument, and 15 questions developed based on existing literature and expert consensus of the coauthors that were specifically focused on understanding the care of older rheumatology patients. The ERA-12 (a simplified version of the ERA-38) is a validated, Likert-scaled, 12-question tool that assesses the strength of stereotypical beliefs toward and expectations regarding aging along three subscales (mental health, physical health, and cognitive function).[26,27] Responses to each of the 12 items were ranked on a 4-point ordinal scale (1, definitely true; 2, somewhat true; 3, somewhat false; and 4, definitely false). A summed score (12–48) was calculated, with higher scores indicating lower or less stereotypical beliefs regarding aging. For simplicity of interpretation, we did not transform the ERA-12 scores and used the natural range from 12 to 48 for each item.

Other questions assessed respondents' attitudes toward and experiences with the rheumatologic care of older adults. Questions regarding prescribing practices were only presented to respondents who identified themselves as prescribers. The coauthors iteratively developed questions based on known literature and clinical expertise in geriatric rheumatology. For example, although it is known that older adults with late-onset RA are undertreated with DMARDs as compared to younger adults, the reasons for this are not fully known.[24] Several questions were developed to explore how rheumatologists approach older adults when making decisions about treatment, and how they counsel these patients (eg, focusing more on potential adverse events as opposed to highlighting known positive effects of a treatment). One question assessed familiarity with the Geriatric 5Ms (mind, mobility, medications [includes polypharmacy], multicomplexity,

and what matters most to the individual), an approach to age-friendly care developed by the Institute for Healthcare Improvement and the American Geriatrics Society, to provide a simple framework that encourages HCPs to focus on principles that would lead to high-value and evidence-based care of older adults.[28,29]

During the 12-month data collection period, links to this survey were distributed via social media (X/Twitter, Facebook, LinkedIn), email, and at rheumatology-related presentations and conferences including the 2023 American College of Rheumatology (ACR) annual scientific meeting. Those who completed the survey were encouraged to share the link to the survey with other individuals in their social and professional network who were also eligible for the study.

Analysis. The survey included 10 Likert-scaled questions, 2 "yes" or "no" questions, and 3 multiple response or free response questions pertinent to the rheumatologic care of older adults and the ERA-12 instrument. Descriptive statistics using medians and proportions were used to report respondent characteristics and their responses. Spearman's rank correlation coefficient was used to assess for correlations between the overall ERA-12 score (a composite of ordinal variables scored 12–48) and responses to each of the Likert-scaled clinical questions and demographic categories. The Mann-Whitney U-test was used to assess for correlations between ERA-12 scores and the responses to the "yes" or "no" questions and two category demographic questions. The Kruskal-Wallis test was used to assess for correlations between ERA-12 scores and demographic categories that were neither Likert scaled nor included only two options. For non–Likert-scaled demographic categories, only categories with five or more participants were included in the statistical analysis. Only data from completed surveys were used in our analysis. Answers to the free response questions were carefully reviewed by APS, PMA, and UEM and categorized based on thematic content. Analyses were done using SPSS.

RESULTS

We collected 255 completed survey responses between January 2023 and December 2023, with 78% of consenting HCPs completing the survey. Survey respondents were predominantly female (63%), White (70%), physicians (75%), practicing in academia (66%) (Table 1). Median time in practice was between 11 and 20 years. A majority of respondents (64%) practiced in an urban setting. Nearly half of respondents' practices (49%) included 25% to 50% patients over the age of 65 years. Responses were received from around the world, with most respondents (82%) being from the United States (Table 1).

The median ERA-12 score was 36 of 48 (interquartile range 8), indicating that respondents, on average, disagreed with the stereotypes regarding aging (Table 2). Female sex ($P = 0.003$) and rural practice setting ($P = 0.027$) were associated with higher ERA-12 scores (suggestive of less stereotypical beliefs regarding aging). Otherwise, no significant interactions were noted between demographic factors and ERA-12 scores (Table 3).

Respondents with lower ERA-12 scores (suggestive of increased stereotypical beliefs regarding aging) were more likely to believe that older adults with rheumatic disease have a negative view of the aging process ($P = 0.004$), are more demanding of attention than younger adults ($P < 0.001$), and are less likely to enjoy caring for older adults ($P < 0.001$). Respondents' perceptions of the beliefs and experiences of older adults with rheumatic disease were also correlated with ERA-12 scores. Lower ERA-12 scores were correlated with belief that older adults are more concerned about the risks of aggressive medical intervention than the possible benefits ($P = 0.018$) and that older adults experience rheumatic disease differently than younger adults ($P = 0.029$). ERA-12 scores were not significantly associated with believing that older adults are more likely to minimize their symptoms than younger adults or reporting that other HCPs express negative attitudes toward older adults (Table 4).

ERA-12 scores were also associated with reported potential changes to the management and counseling of older adults. Lower ERA-12 scores were associated with changing the focus from disease-modifying therapy to symptom relief when caring for older adults ($P < 0.001$) and emphasizing the risks of aggressive medical intervention to older adults more than younger adults ($P = 0.006$). ERA-12 scores were not significantly associated with the respondent's confidence in caring for older adults (Table 4).

Our questions assessing the effect of training specific to the care of older adults were mixed in their findings. Nearly 60% of the respondents were not familiar with the Geriatric 5Ms and had not received specific training in the care of older adults. HCPs familiar with the Geriatric 5Ms (41%) were more likely to have higher ERA-12 scores ($P < 0.001$), suggesting lower ageist stereotypes. There was no significant correlation between the reported history of training specific to the care of older adults and ERA-12 scores (Table 4).

The most frequently identified challenging aspects of caring for older adults included multicomplexity (91%), polypharmacy (77%), and lack of time to address issues during an office visit (52%) (Table 5). Respondents with higher ERA-12 scores were significantly more likely to report having insufficient time to address patient issues during an office visit. Free responses were categorized by theme, with commonly described challenges including transportation problems, insufficient technologic literacy, lack of sufficient caregiver support, care fragmentation, and financial constraints.

The most frequently identified ways respondents treat older adults differently than younger adults included decreased use of nonsteroidal anti-inflammatory drugs (NSAIDs) (79%) and opioids (43%) in older adults. One in five respondents reported they treated older adults the same way they treated younger adults

Table 1. Demographics of survey respondents (N = 255)*

Demographic	Survey Respondents, n (%)
Age, y	
<30	28 (11)
30–44	91 (36)
45–64	115 (45)
>65	21 (8)
Gender	
Woman	160 (63)
Man	94 (37)
Prefer not to say	1 (0)
Race and ethnicity	
American Indian or Alaskan Native	3 (1)
Asian	58 (23)
Black or African American	3 (1)
White	178 (70)
Other	9 (4)
Prefer not to say	7 (3)
Hispanic	
Yes	23 (9)
No	231 (91)
Prefer not to say	1 (0)
Continent or country of residence	
Africa	0 (0)
Asia	9 (4)
Europe	11 (4)
Oceania	10 (4)
South America	2 (1)
United States of America	210 (82)
Other North America	13 (5)
Specialty	
Rheumatology	185 (84)
Internal medicine	28 (13)
Other	7 (3)
Occupation	
Advanced practice provider	29 (11)
Nurse	3 (1)
Pharmacist	7 (3)
Physician	191 (75)
Psychologist	0 (0)
Rehabilitation (PT/OT)	10 (4)
Social worker	1 (0)
Chiropractor	8 (3)
Other	6 (2)
Years in practice	
In training	44 (17)
<10 y	64 (25)
11–20 y	76 (30)
21–30 y	48 (19)
>30 y	22 (9)
Prefer not to say	1 (0)
Practice setting	
Academic medicine	168 (66)
Private practice	50 (20)
Large group practice	23 (9)
Government	14 (6)
Patient community	
Urban	164 (64)
Suburban	72 (28)
Rural	19 (8)
Percent of practice age >65 y	
Less than 25%	31 (12)
25–50%	124 (49)
51–75%	94 (37)
Greater than 75%	6 (2)

Table 1. (*Cont'd*)

* OT, occupational therapy; PT, physical therapy.

(Table 5). None of the reported changes to the treatment of older adults were significantly associated with ERA-12 scores. Free responses were categorized by theme, with common responses including increased focus on comorbidities when choosing medications, a decreased dose of DMARDs in older adults, and medication changes due to financial constraints.

The most frequently chosen resources to improve the care of older adults were geriatrics consultation and collaborative care (64%), physical and occupational therapy (58%), more time or more frequent visits (48%), and pharmacy consultation (41%) (Table 5). Free responses were limited, and no pattern could be discerned.

DISCUSSION

This is the first study, to our knowledge, that evaluates the HCP's perspective on caring for older adults in rheumatology. We dive into understanding attitudes and stereotypes regarding age (ageism) and how this may impact HCP's perception and interaction with older adults with rheumatic diseases. In this survey-based study, we found that increased stereotypical thinking regarding aging is associated with changes to HCPs' perceptions of the care of older adults and to their perception of the aging experience of their older adult patients.

Rheumatology HCPs with lower ERA-12 scores (increased stereotypical beliefs regarding aging) were more likely to believe that older adults have a negative view of the aging process, are more demanding of attention, are more concerned about the risks of medical interventions (rather than the possible benefits), and experience rheumatic disease differently than younger adults. It is not possible to identify the direction of causation, whether experiences with caring for older adults have led to strengthening of stereotypical beliefs regarding aging or, conversely, that underlying stereotypical beliefs are worsening the provider's experience caring for older adults. Conscious or unconscious attitudes toward the aging population may lead to assumptions and may shift the approach to communicating with and managing older adults with rheumatic diseases. Research in other fields has linked bias regarding aging to quantifiable changes to patient counseling and treatment decisions.[30] Indeed, we found an association between increased stereotypical beliefs regarding aging and potential

Table 2. ERA-12 instrument responses by question (N = 255)*

	Definitely true, n (%)	Somewhat true, n (%)	Somewhat false, n (%)	Definitely false, n (%)	Median (IQR)
Physical health subscale					11 (3)
When people get older, they need to lower their expectations of how healthy they can be.	8 (3)	75 (29)	78 (31)	94 (37)	3
The human body is like a car: when it gets old, it gets worn out.	21 (8)	139 (55)	74 (29)	21 (8)	2
Having more aches and pains is an accepted part of aging.	17 (7)	119 (47)	83 (33)	36 (14)	2
Every year that people age, their energy levels go down a little more.	9 (4)	86 (34)	124 (49)	36 (14)	3
Mental health subscale					15 (3)
I expect that as I get older, I will spend less time with friends and family.	3 (1)	22 (9)	61 (24)	169 (66)	4
Being lonely is just something that happens when people get old.	0 (0)	27 (11)	45 (18)	183 (72)	4
Quality of life declines as people age.	1 (0)	73 (29)	82 (32)	99 (39)	3
It's normal to be depressed when you are old.	0 (0)	6 (2)	50 (20)	199 (78)	4
Cognitive function subscale					11 (4)
I expect that as I get older, I will become more forgetful.	14 (5)	112 (44)	86 (34)	43 (17)	3
It's an accepted part of aging to have trouble remembering names.	9 (4)	95 (37)	92 (36)	59 (23)	3
Forgetfulness is a natural occurrence just from growing old.	5 (2)	70 (27)	112 (44)	68 (27)	3
It is impossible to escape the mental slowness that happens with aging.	7 (3)	59 (23)	111 (44)	78 (31)	3
Overall ERA-12 score					36 (8)

* For calculation of medians: definitely true = 1, somewhat true = 2, somewhat false = 3, and definitely false = 4.
ERA-12, Expectations Regarding Aging; IQR, interquartile range.

changes to patient counseling and medical decision-making. Specifically, providers with increased stereotypical beliefs regarding aging were more likely to emphasize the risks of aggressive medical intervention to older adults than younger adults and were more likely to focus on symptom relief than disease-modifying therapy. This suggests that the care older adults are offered and choose to receive may vary depending on the degree of age-associated bias of their HCP. This may in part explain the reduced use of biologics and targeted synthetic DMARDs in older adults with RA.[18–24]

The degree to which communication and counseling is modified by ageist beliefs is not known. However, it is plausible that changes to patient counseling could contribute to the belief that older adults are more concerned about the risks of aggressive medical intervention (rather than the possible benefits). Older adults are more likely to have conditions that may make certain medication choices more risky or less appropriate including renal disease, liver disease, recurrent infection(s), frailty, and polypharmacy.[31] However, the older adult population is heterogeneous, and not every older adult has all or even any of these conditions. Therefore, we must be vigilant that underlying biases regarding age do not lead to changes in practice patterns based on age alone rather than the patient in front of us. Importantly, when decisions to offer (or not offer) older adults certain medications are based on chronological age alone, rather than the current and desired functional status, we run the risk of undertreatment, which may fall short of the individual's preferences, priorities, and functional goals.[32]

Our results suggest that respondents who were women or practicing in rural settings had higher ERA-12 scores (indicative of lower ageist stereotypes). The reasons for this are unclear, but other studies have similarly shown that American women, including women physicians, have higher ERA-12 scores than men.[33] Some suggested possible reasons include differences in communication styles, variations in clinical training, and attitudes regarding the patient–physician relationship, but the underlying reasons for this have not been rigorously explored. Regarding the significant association for rural respondents, we would recommend interpreting this finding with caution because the overall sample size for this group of respondents was limited.

Rheumatologists with reduced stereotypical beliefs regarding aging were more likely to report that they enjoy caring for older adults and are aware of the Geriatric 5Ms. The Geriatric 5Ms are a popular framework used to help guide appropriate care of older adults.[34] The positive correlation between awareness of these principles and the level of stereotypical beliefs regarding aging suggests that education of rheumatology professionals in geriatric principles could reduce stereotypical beliefs regarding age and improve provider comfort, awareness, and competence and as

Table 3. Demographic factors and ERA-12 correlations (N = 255)*

Demographic factor	Respondents, n (%), median ERA-12	Summary statistic	*P* value
Age, y		−0.08	ns
<30	28 (11), 36.5		
30–44	91 (36), 37		
45–64	115 (45), 35		
>65	21 (8), 35		
Gender		2.96	0.003
Woman	160 (63), 37		
Man	94 (37), 35		
Hispanic		−1.00	ns
Yes	23 (9), 37		
No	231 (91), 36		
Race		1.37	ns
Asian	58 (23), 35.5		
White	178 (70), 37		
Years in practice		−0.08	ns
In training	44 (17), 36.5		
<10 y	64 (25), 37		
11–20 y	76 (30), 37		
21–30 y	48 (19), 35		
>30 y	22 (9), 35		
Practice setting		2.31	ns
Academics	168 (66), 36		
Private practice	50 (20), 36		
Government	14 (6), 37.5		
Large group practice	23 (9), 35		
Patient community		7.23	0.027
Urban	164 (64), 36		
Suburban	72 (28), 35		
Rural	19 (8), 39		
Percent of practice age >65 y		0.08	ns
<25%	31 (12), 35		
25–50%	124 (49), 35.5		
51–75%	94 (37), 37		
>75%	6 (2), 43.5		

* Categories with <5 participants excluded from statistical analysis.
ERA-12, Expectations Regarding Aging; ns, not significant.

a result potentially reduce the risk of age-related bias affecting the care of older adults.[35]

The responses to our free response and multiple response questions highlighted multiple key geriatric principles, which is perhaps unsurprising, as ERA-12 scores on average were quite high in our sample. Several of the 5Ms were identified as being particularly challenging aspects of caring for older adults with rheumatic diseases, including multicomplexity, polypharmacy, and cognitive impairment. The most commonly reported prescribing change when caring for older adults was reduced prescribing of NSAIDs and opioids, which is perhaps appropriate because these medications are listed as potentially inappropriate medications in the Beers criteria.[31]

The need for an interdisciplinary team (geriatrics or collaborative care, physical or occupational therapy) was the most important provider identified element to improve the care of older adults. This aligns well with the ACR 2022 Clinical Practice Guidelines for the integrative management of RA,[36] which stress the key role of physical and occupational therapy in caring for this patient group. Respondents highlighted the need for more time and frequent visits to improve care for older adults. Recent regulatory changes around outpatient billing allowing for the use of time- and complexity-based billing may help to improve the care of medically complex older adults. Further changes to policy and reimbursement rates, along with innovative models of health care delivery, may be necessary to further improve access, equity, and interdisciplinary management of older adults with rheumatic disease.[37]

There are several limitations to this study related to survey administration. In this study, we attempted to recruit broadly and capture diverse perspectives on caring for older adults in rheumatology. However, recruitment of participants was dependent on professional networks and snowball recruitment, biasing the sample toward academic rheumatologists. The academic workforce is an important target for intervention due to the disproportionately high concentration of medically complex older adults and the Medicare population in academic clinics.[38] Our outreach and dissemination strategy was most effective when we could identify a champion to help reach and encourage their colleagues to complete the survey. This was most critical for the recruitment of non-US and nonphysician respondents. Recruitment via social media, including X/Twitter, and having the survey title read "Ageism in Rheumatology: The Clinical Team's Perspective" likely enriched the sample for a population of rheumatologists interested in the care of older adults, which may explain the high average ERA-12 scores in this study. Survey respondents were predominantly White and Asian, with very few responses from Black or Hispanic colleagues, similar to findings seen in the 2015 ACR workforce study.[39] Our respondents were younger on average than those in the 2015 ACR workforce study, which may reflect our social media–based recruitment strategy. Our respondents were also more likely to be female, which mirrors the ongoing shift from a male-predominant to a female-predominant rheumatology workforce.[39] An alternative recruiting approach will be needed in future studies seeking to gather perspectives from a broader interprofessional, international sample. Survey administration via a link required internet availability, further limiting reach and generalizability. We acknowledge that this correlative analysis is not able to confirm if certain beliefs about aging correspond directly to specific physician behaviors or treatment changes when caring for older adults, and, indeed, the rheumatology-specific questions were created for this survey and have not been validated.

Our findings speak to the importance of identifying age-related biases in the field of rheumatology to ensure older adults receive the standard of care. Future studies are needed to extend the scope of this study to assess the magnitude of bias affecting the care of patients with rheumatic diseases and resulting

Table 4. Likert-scaled and "yes" or "no" questions with ERA-12 correlations (N = 255)*

	Strongly agree, n (%), median ERA-12	Agree, n (%), median ERA-12	Disagree, n (%), median ERA-12	Strongly disagree, n (%), median ERA-12	Yes, n (%), median ERA-12	No, n (%), median ERA-12	Spearman ρ[a] or U statistic[b]	*P* value
I am confident managing the care of older adults with rheumatic disease.	88 (35), 37	**139 (55), 37**	26 (10), 33	2 (1), 34	–	–	−0.095[a]	ns
Older adults experience rheumatic diseases differently than younger adults.	22 (9), 35	**144 (57), 36**	83 (33), 37	6 (2), 44	–	–	0.137[a]	0.029
Older adults with rheumatic disease have a negative attitude toward the aging process.	6 (2), 40.5	63 (25), 35	**163 (64), 37**	23 (9), 40	–	–	0.179[a]	0.004
Other health care professionals express negative attitudes about older people.	6 (2), 40.5	52 (20), 36	**155 (61), 36**	42 (17), 36	–	–	0.032[a]	ns
Older adults are more likely to minimize their symptoms than younger adults.	19 (8), 36	**119 (47), 37**	106 (42), 35.5	11 (4), 35	–	–	−0.014[a]	ns
Older adults are more demanding of attention than younger adults.	1 (0), 38	24 (9), 33	**174 (68), 36**	56 (22), 38	–	–	0.239[a]	<0.001
One way I treat older adults differently than younger adults is that I tend to focus my prescribing efforts more on symptom relief than disease-modifying therapy.[c]	1 (0), 30	45 (21), 35	**124 (56), 35.5**	50 (23), 39	–	–	0.233[a]	<0.001
One way I treat older adults differently than younger adults is that I tend to emphasize the risks of aggressive medical intervention rather than the possible benefits.[c]	2 (1), 39	67 (31), 35	**119 (54), 37**	32 (15), 39	–	–	0.184[a]	0.006
Older adults are typically more concerned about the risks of aggressive medical intervention than they are interested in the possible benefits.	10 (4), 34	94 (37), 35	**130 (51), 37**	21 (8), 38	–	–	0.148[a]	0.018
I enjoy caring for older adults with rheumatic disease.	103 (40), 38	**141 (55), 35**	11 (4), 38	0 (0)	–	–	−0.254[a]	<0.001
Are you familiar with the Geriatric 5Ms?	–	–	–	–	104 (41), 38	**151 (59), 35**	−3.601[b]	<0.001
Have you received any training specific to the care of older adults?	–	–	–	–	103 (40), 37	**152 (60), 35**	−1.575[b]	ns

* The most frequent responses to each question are bolded. ERA-12, Expectations Regarding Aging; ns, not significant.
[a] Indicates Spearman's ρ.
[b] Indicates U statistic.
[c] n = 220 for questions presented only to prescribers.

outcomes. There are several potential targets for interventions to combat ageism, including those outlined in the World Health Organization Global Campaign to Combat Ageism.[40] Our findings suggest that knowledge of aging principles is associated with reduced stereotypical beliefs regarding aging. Additionally, the literature suggests that interventions during medical training may modestly reduce the strength of age-related stereotypes in trainees,[41] although the durability of these changes and their

Table 5. Multiple response questions*

	Yes, n (%)	Median ERA-12		z score	*P* value
		Yes	No		
What are some of the ways you treat older adults with rheumatic disease differently than younger adults? (n = 220)					
Decreased use of NSAIDs	174 (79)	37	36	−0.02	ns
Decreased use of opioids	95 (43)	35	37	1.14	ns
I manage older adults the same way	39 (18)	36	37	−0.62	ns
Decreased use of biologics	38 (17)	37	36	0.40	ns
Increased use of oral glucocorticoids	37 (17)	35	37	1.59	ns
Decreased use of DMARDs	31 (14)	33	37	1.92	ns
Decreased use of intra-articular steroids	2 (1)	28.5	36	na	na
Other	16 (7)	–	–	na	na
What are the most challenging aspects of caring for older adults in your clinic? (n = 255)					
They have multiple medical conditions (multicomplexity)	232 (91)	36	37	−0.27	ns
They have long medication lists (polypharmacy)	196 (77)	36	36	−0.20	ns
There is not enough time to address issues during an office visit	133 (52)	37	35	2.44	0.015
They are financially burdened or limited	112 (44)	36.5	36	−0.56	ns
Insufficient data to guide medical decisions	75 (29)	37	35.5	−1.05	ns
They can't access medications	72 (28)	37	36	−1.61	ns
They require more resources	71 (28)	36	36	0.06	ns
They have impaired hearing and/or vision	71 (28)	37	36	0.44	ns
They have cognitive impairment	68 (27)	35	37	0.95	ns
They have poor health literacy	39 (15)	37	36	−0.19	ns
They have poor insight into their medical care	35 (14)	34	37	1.12	ns
They want to manage their own treatment	10 (4)	32.5	37	1.17	ns
They are emotionally draining	8 (3)	32.5	36	0.96	ns
They are demanding or attention-seeking and complain	6 (2)	31.5	36	1.20	ns
They have offensive behavior or are thankless	0 (0)	na	36	na	na
What are the top three resources that can best help you improve your care of older adults with rheumatic disease? (n = 255)					
Geriatrics consultation or collaborative care	162 (64)	–	–	–	–
Physical and occupational therapy	147 (58)	–	–	–	–
More time or frequent visits	133 (52)	–	–	–	–
Pharmacy	104 (41)	–	–	–	–
Social work	87 (34)	–	–	–	–
Clinical psychology	52 (20)	–	–	–	–
Decision aids	40 (16)	–	–	–	–
Other	8 (3)	–	–	–	–

* DMARD, disease-modifying antirheumatic drug; ERA-12, Expectations Regarding Aging; na, not available; ns, not significant; NSAID, nonsteroidal anti-inflammatory drug.

effect on patient care are unclear. New tools and curricula to improve education in geriatric principles could be helpful in future interventions to reduce the effects of ageism on older adult patients.[40,42] Importantly, the patient perspective on the prevalence and effects of ageism on rheumatologic care needs to be assessed.[43,44] Although we may not currently have the means to measure precisely how ageism impacts the geriatric population in rheumatology, it is important to reflect on our own potential prejudices early on and strive to deliver patient care that is free from bias.

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 Makris 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. United States Census Bureau. 65 and Older Population Grows Rapidly as Baby Boomers Age. US Dept of Commerce; 2020. Accessed August 10, 2024. https://www.census.gov/newsroom/press-releases/2020/65-older-population-grows.html
2. Cleutjens FAHM, Boonen AERCH, van Onna MGB. Geriatric syndromes in patients with rheumatoid arthritis: a literature overview. Clin Exp Rheumatol 2019;37(3):496–501.
3. Levinson J, Buehring B. Pitfalls and pearls in diagnosing inflammatory arthritis in older patients. Joint Bone Spine 2024;91(6):105719.

4. 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.
5. Gunderson TM, Myasoedova E, Davis JM III, et al. Multimorbidity burden in rheumatoid arthritis: a population-based cohort study. J Rheumatol 2021;48(11):1648–1654.
6. Boots AMH, Maier AB, Stinissen P, et al. The influence of ageing on the development and management of rheumatoid arthritis. Nat Rev Rheumatol 2013;9(10):604–613.
7. van Onna M, Boonen A. Challenges in the management of older patients with inflammatory rheumatic diseases. Nat Rev Rheumatol 2022;18(6):326–334.
8. Buehring B, van Onna M, Myasoedova E, et al. Understanding the multiple dimensions of ageing: 5Ms for the rheumatologist. Lancet Rheumatol 2024;6(12):e892–e902.
9. World Health Organization. *Global Report on Ageism*. 2021. Accessed August 10, 2024. https://iris.who.int/handle/10665/340208
10. Chang ES, Kannoth S, Levy S, et al. Global reach of ageism on older persons' health: a systematic review. PLoS One 2020;15(1):e0220857.
11. Levy BR, Slade MD, Chang ES, et al. Ageism amplifies cost and prevalence of health conditions. Gerontologist 2020;60(1):174–181.
12. Allen JO, Solway E, Kirch M, et al. Experiences of everyday ageism and the health of older US Adults. JAMA Netw Open 2022;5(6):e2217240.
13. Levy BR, Myers LM. Preventive health behaviors influenced by self-perceptions of aging. Prev Med 2004;39(3):625–629.
14. Kim ES, Moored KD, Giasson HL, et al. Satisfaction with aging and use of preventive health services. Prev Med 2014;69:176–180.
15. Liu YE, While AE, Norman IJ, et al. Health professionals' attitudes toward older people and older patients: a systematic review. J Interprof Care 2012;26(5):397–409.
16. Fraenkel L, Rabidou N, Dhar R. Are rheumatologists' treatment decisions influenced by patients' age? Rheumatol Oxf Engl 2006;45(12):1555–1557.
17. Kievit W, van Hulst L, van Riel P, et al. Factors that influence rheumatologists' decisions to escalate care in rheumatoid arthritis: results from a choice-based conjoint analysis. Arthritis Care Res (Hoboken) 2010;62(6):842–847.
18. Vontela S, Haq Z, Joseph V, et al. Age-related differences in the treat-to-target approach to rheumatoid arthritis management in an urban clinic. J Clin Rheumatol 2022;28(6):321–324.
19. Tutuncu Z, Reed G, Kremer J, et al. Do patients with older-onset rheumatoid arthritis receive less aggressive treatment? Ann Rheum Dis 2006;65(9):1226–1229.
20. Nilsson J, Andersson MLE, Hafström I, et al. Influence of age and sex on disease course and treatment in rheumatoid arthritis. Open Access Rheumatol 2021;13:123–138.
21. Steffen A, Holstiege J, Klimke K, et al. Patterns of the initiation of disease-modifying antirheumatic drugs in incident rheumatoid arthritis: a German perspective based on nationwide ambulatory drug prescription data. Rheumatol Int 2018;38(11):2111–2120.
22. Jung SM, Kwok SK, Ju JH, et al. Risk factors associated with inadequate control of disease activity in elderly patients with rheumatoid arthritis: results from a nationwide Korean College of Rheumatology Biologics (KOBIO) registry. PLoS One 2018;13(10):e0205651.
23. DeWitt EM, Lin L, Glick HA, et al. Pattern and predictors of the initiation of biologic agents for the treatment of rheumatoid arthritis in the United States: an analysis using a large observational data bank. Clin Ther 2009;31(8):1871–1880; discussion 1858.
24. Lee J, Martindale J, Makris UE, et al. Initiation of disease-modifying antirheumatic drugs in older Medicare beneficiaries with new diagnosis of late-onset rheumatoid arthritis. ACR Open Rheumatol 2023;5(12):694–700.
25. Radovits BJ, Fransen J, Eijsbouts A, et al. Missed opportunities in the treatment of elderly patients with rheumatoid arthritis. Rheumatology (Oxford) 2009;48(8):906–910.
26. Sarkisian CA, Steers WN, Hays RD, et al. Development of the 12-item expectations regarding aging survey. Gerontologist 2005;45(2):240–248.
27. Ayalon L, Dolberg P, Mikulionienė S, et al. A systematic review of existing ageism scales. Ageing Res Rev 2019;54;100919.
28. Tinetti M, Huang A, Molnar F. The Geriatrics 5M's: a new way of communicating what we do. J Am Geriatr Soc 2017;65(9):2115.
29. Institute for Healthcare Improvement. Age-Friendly Health Systems. Accessed August 10, 2024. https://www.ihi.org/networks/initiatives/age-friendly-health-systems
30. Madan AK, Aliabadi-Wahle S, Beech DJ. Age bias: a cause of underutilization of breast conservation treatment. J Cancer Educ 2001;16(1):29–32.
31. By the 2023 American Geriatrics Society Beers Criteria Update Expert Panel. American Geriatrics Society 2023 updated AGS Beers Criteria for potentially inappropriate medication use in older adults. J Am Geriatr Soc 2023;71(7):2052–2081.
32. Fraenkel L, Bogardus ST, Concato J, et al. Patient preferences for treatment of rheumatoid arthritis. Ann Rheum Dis 2004;63(11):1372–1378.
33. Davis MM, Bond LA, Howard A, et al. Primary care clinician expectations regarding aging. Gerontologist 2011;51(6):856–866.
34. Schwartz AW, Hawley CE, Strong JV, et al. A workshop for interprofessional trainees using the Geriatrics 5Ms framework. J Am Geriatr Soc 2020;68(8):1857–1863.
35. Miller RK, Young M, Chippendale R, et al. Using the Geriatric 5Ms to teach structural and social determinants of health. J Am Geriatr Soc 2023;71(12):3967–3972.
36. England BR, Smith BJ, Baker NA, et al. 2022 American College of Rheumatology guideline for Exercise, Rehabilitation, Diet, and Additional Integrative Interventions for Rheumatoid Arthritis. Arthritis Care Res (Hoboken) 2023;75(8):1603–1615.
37. Farrell TW, Hung WW, Unroe KT, et al. Exploring the intersection of structural racism and ageism in healthcare. J Am Geriatr Soc 2022;70(12):3366–3377.
38. Burke L, Khullar D, Orav EJ, et al. Do academic medical centers disproportionately benefit the sickest patients? Health Aff (Milwood) 2018;37(6):864–872.
39. American College of Rheumatology. 2015 Rheumatology Workforce Study of Rheumatology Specialists in the United States. 2016. Accessed April 1, 2025. https://assets.contentstack.io/v3/assets/bltee37abb6b278ab2c/bltb4276e6740a51375/632a340a27a8dd5d8cd3ddc0/rheumatology-workforce-study-2015.pdf
40. World Health Organization. *Global Campaign to Combat Ageism - Toolkit*. 2021. Accessed October 6, 2024. https://www.who.int/publications/m/item/global-campaign-to-combat-ageism-toolkit
41. Morgan S, Pasco JC, Demers L, et al. Combating ageism in medical education with narrative medicine. Gerontol Geriatr Educ 2024;46(1):5–16.
42. Krichbaum K. Addressing ageism in healthcare - overview. Regents of the University of Minnesota. 2022. Accessed October 6, 2024. https://mngwep.nexusipe.org/static/addressing-ageism/index.html#/
43. Lee J, Misra D, Makris UE. Ageism in rheumatologic care: ensuring equity and quality for older adults. Arthritis Care Res (Hoboken) 2025;77(10):1175–1179.
44. Misra D, Buehring B, Yung R, Makris UE. Ageism and rheumatic diseases. Lancet Rheumatol. 2024; 6(12):e817–e819. doi: https://doi.org/10.1016/S2665-9913(24)00239-X. Epub 2024 Nov 11. PMID: 39542000.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1263–1272
DOI 10.1002/acr.25556

American College of Rheumatology
Empowering Rheumatology Professionals

A Consensus-Based Shoulder Examination for Rheumatology Training

H. Berk Degirmenci,[1] Robert Kalish,[2] Ratnesh Chopra,[3] Katarzyna Gilek-Seibert,[4] Fotios Koumpouras,[5] Vasileios Kyttaris,[6] Santhanam Lakshminarayanan,[7] Bonita Libman,[8] Stephanie Mathew,[9] Kenneth O'Rourke,[10] Anthony M. Reginato,[11] and Eugene Kissin[1]

Objective. Physical examination of a patient with nonspecific shoulder pain is a nonstandardized teaching objective among rheumatology fellowship programs. We investigated consensus among rheumatology fellowship program directors (PDs) in shoulder examination maneuvers to be performed by rheumatology fellows.

Methods. Past or present rheumatology PDs currently working in New England taught a five-minute shoulder examination to one of their trainees and recorded the resulting video of the shoulder examination. We cataloged all the performed maneuvers from these videos. Anonymized electronic surveys instructed PDs to rank each maneuver into one of the three categories: teach all fellows and should be performed routinely (tier 1); teach all fellows but should be performed only in specific scenarios (tier 2); and each only to selected fellows and should be performed only in specific scenarios (tier 3). For maneuvers performed differently, we surveyed for consensus. Items not meeting the ≥70% consensus threshold were included in the second survey, and this process was repeated for a third survey. A separate survey collected PD demographics.

Results. Eleven of 13 recruited PDs agreed to participate, and 100% of participants completed all rounds of the study. The study addressed 65 items: 52 questions for tier designation of the examination maneuvers and 13 questions for different examination techniques. Participants achieved consensus for 40 of 52 tier designation items and for 8 of 13 technique items.

Conclusion. This is the first study focused on shoulder examination specific to rheumatology practice, and these results can provide high-yield guidance for the rheumatology community.

INTRODUCTION

Shoulder pain is a common cause of morbidity in the general population and is the third most common cause of musculoskeletal consultation in primary care.[1,2] A US-based study from general internists reported that 40% of patients with shoulder pain are referred to orthopedic surgeons or rheumatologists.[3] Rotator cuff pathologies comprise the most common underlying cause of shoulder pain, followed by periarticular soft tissue pain and impingement syndromes in community-based adult rheumatology practice and primary care in the United Kingdom.[4,5]

American College of Rheumatology (ACR) learning programs "Rheum2Learn" and "Virtual Rheumatology Program for Fellows in Training" aid trainees in learning shoulder examination. ACR lecturers emphasize that shoulder examination in rheumatology is not standardized given complex regional anatomy, limitations of the targeted maneuvers, and different learning experiences during fellowship training that are carried over to postgraduation practice.[6] Hereby, we propose a standardized algorithm for shoulder examination for those who teach rheumatology trainees to consider in their own pedagogic

[1]Berk Degirmenci, MD, Eugene Kissin, MD: Boston University Chobanian and Avedisian School of Medicine, Boston, Massachusetts; [2]Robert Kalish, MD: Tufts University School of Medicine, Boston, Massachusetts; [3]Ratnesh Chopra, MD: University of Massachusetts Chan Medical School, Worcester; [4]Katarzyna Gilek-Seibert, MD: Roger Williams Medical Center, Providence, Rhode Island; [5]Fotios Koumpouras, MD: Yale School of Medicine, New Haven, Connecticut; [6]Vasileios Kyttaris, MD: Beth Israel Deaconess Medical Center, Boston, Massachusetts; [7]Santhanam Lakshminarayanan, MD: University of Connecticut School of Medicine, Farmington; [8]Bonita Libman, MD: The University of Vermont Medical Center, Burlington; [9]Stephanie Mathew, DO: Dartmouth-Hitchcock Medical Center, Hanover, New Hampshire; [10]Kenneth O'Rourke, MD: Maine Medical Center, Portland; [11]Anthony M. Reginato, MD, PhD: The Warren Alpert Medical School of Brown University, Providence, Rhode Island.

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.25556).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25556.

Address correspondence via email to Eugene Kissin, MD, at eukissin@bu.edu.

Submitted for publication August 22, 2024; accepted in revised form March 11, 2025.

SIGNIFICANCE & INNOVATIONS

- We created a list of shoulder examination maneuvers recommended for all rheumatology trainees to learn and perform when evaluating patients with nondescript shoulder pain.
- Delphi methodology–based anonymized survey rounds allowed an expert panel to review maneuvers and techniques used in rheumatology training programs, with subsequent assignment to tier designations that can provide high-yield information for medical educators in the rheumatology community.
- The addition of short video clips for items assessed variations in examination techniques minimized potential misinterpretation of different techniques and allowed experts to provide clear insights.
- This study is most directly relevant to the rheumatology community, but we believe it may also serve as a useful methodologic guidance document for other educators in the musculoskeletal diseases field who are interested in studying challenging yet important components of the musculoskeletal examination.

settings and to assist in developing a consistent learning strategy for trainees.

PATIENTS AND METHODS

Participants. We recruited current or past rheumatology fellowship program directors (PDs) from New England to form an expert panel for this study. Each member of the panel recruited a trainee (rheumatology fellow or internal medicine resident) to whom they would teach the shoulder examination. Shoulder model volunteers were recruited from house staff.

Data collection. The panel was asked to teach their trainees a five-minute physical examination of the shoulder for use in a patient with nonspecific shoulder pain, when there are no additional historical clues, thus stimulating the expert panel to create a comprehensive yet concise list of maneuvers for the rheumatology setting. Subsequently, experts recorded the shoulder examinations performed by their trainees on model volunteers and uploaded resulting videos to a shared online database.

The lead author and corresponding author (HBD and EK), reviewed all the resulting videos and cataloged all the performed maneuvers and all the techniques for each maneuver. Anonymity was fully preserved for all panel members except the corresponding author, for which anonymity was partial due to recognition of trainees in 3 of 11 resulting videos. Comprehensive listing of the shoulder examination was subdivided into five different domains (inspection, palpation, active range of motion [ROM], passive ROM, and resistance testing). The final content was used for this Delphi-based survey consisting of three rounds.

Study protocol. Electronic anonymized surveys were created using "Qualtrics" (Qualtrics, LLC). Anonymous surveys collected the characteristics of the expert panel. A predetermined threshold of ≥70% was selected to establish consensuses for all survey rounds.

The first-round survey instructed the expert panel to rank each maneuver into one of the three tiers: teach all fellows and should be performed routinely (tier 1); teach all fellows but should be performed only in specific scenarios (tier 2); and teach only to selected fellows and should be performed only in specific scenarios (tier 3). We encouraged the panel to include comments to explain their reasoning and provide literature support for their choices. The panel received an anonymized compilation of comments and results of the first survey round. Items not meeting consensus were included in the second survey, and the panel repeated the process. Compiled comments and results were again shared with the panel following the second round, and items not meeting consensus were included in the third (final) round. Experts were encouraged to review anonymized comments and the cited literature between survey rounds to increase answer malleability.

These surveys also included items listing different techniques for maneuvers performed variably, along with video clips of the different techniques extracted from submitted videos. These video clips cropped out the face of the trainee or model volunteer to maintain anonymity and reduce the bias for panel voting. The panel was asked to identify their choice for the best technique among the ones performed differently.

Statistical analysis. We used Qualtrics Analysis by Qualtrics, LLC, an embedded basic analysis software for Qualtrics surveys that provides descriptive information of the data obtained from survey answers. Results from the demographics survey were reported as means ± SDs and range format. Survey rounds assessed a predetermined consensus threshold of ≥70% following Delphi methodology.

Video demonstrations and ethics. Demonstrations of the tier 1 and tier 2 maneuvers in the Supplementary Material included the techniques for which consensus was achieved. This study was reviewed and approved by the Boston Medical Center and Boston University Medical Campus Institutional Review Board.

RESULTS

Characteristics of the expert panel. Of 13 potential participants recruited, 10 current and 1 former rheumatology fellowship PDs agreed to participate. Ten current rheumatology

fellowship PDs comprise 83% of the current PDs in the New England area.[7] A total of 100% of participants completed the anonymous demographics surveys and all three rounds of the Delphi exercise. The mean ± SD for the number of years worked as a clinical rheumatologist was 21 ± 7.7 years, with 9 ± 5.6 years of their career spent as a fellowship PD. PDs reported that 47% of their shoulder examination skills stem from their own fellowship training experience. PD training locations included institutions in the Northeast, South, Midwest, and West, and none of the PDs were trained at the same fellowship program (Table 1).

Tier designation of the examination maneuvers, technique-related items, and inspection. The study addressed 65 items: 52 questions for tier designation of the examination maneuvers and 13 questions for different examination techniques. Consensus for tier designation items increased from 44% in round 1 to 69% in round 2 and 77% by round 3. The panel achieved consensus for 40 of 52 tier designation items (Table 2) and for 8 of 13 technique-related items (Table 3). For technique-related items, consensus increased from 31% in round 1 to 54% in round 2 and 61% by round 3 (Figure 1).

The final list for tier 1 maneuvers included three elements of inspection, seven of palpation, four active and four passive ROM maneuvers, and seven resistance testing maneuvers (Table 2). The final list for tier 2 maneuvers included three palpation maneuvers, five active ROM maneuvers, and seven resistance testing maneuvers (Table 2). The demonstrations of tier 1 and tier 2 maneuvers, performed by a rheumatology fellow in training (HBD), included in the Supplementary Material (Supplementary Videos 1 and 2).

No maneuvers met the consensus threshold for tier 3. Twelve maneuvers did not meet the consensus threshold for any tier. This heterogeneous group included maneuvers that were close to meeting the consensus threshold for tier 1 or tier 2 or maneuvers for which the majority felt tier 3 was the most reasonable (Table 4). All inspection maneuvers performed by examiners met the consensus threshold for tier 1 during the first survey round, highlighting the importance adequate exposure of both shoulders with careful observation from the front and back to evaluate for any asymmetry, misalignment, muscle atrophy, and swelling of the articular and nonarticular structures of the shoulder.

Palpation. *Tier 1 palpation maneuvers.* Palpation of the sternoclavicular joint, acromioclavicular joint, acromion, subacromial space, glenohumeral joint, and long head of the biceps tendon met the consensus threshold for tier 1. Palpation of clavicle, as a transition from sternoclavicular joint to acromioclavicular joint, was added to tier 1 during survey round 3.

Palpation of the biceps long head was performed by two different methods: static humeral head or dynamic humeral head (internal and external rotation during palpation). The majority of the experts voted for the dynamic humeral head as the preferred method. Palpation of the glenohumeral joint was also accomplished by two different methods: anterior palpation only or

Table 1. Demographics of the expert panel*

Demographic	Result
1A. Quantitative background information, mean ± SD (range)	
Number of years working as clinical rheumatologist	21 ± 7.7 (9–31)
Number of years serving (or had served) as fellowship program director	9 ± 5.6 (3–22)
Number of fellows accepted to their rheumatology fellowship program (per year)	2 ± 0.6 (1–3)
Estimated percentage of MSK examination knowledge fellows learn(ed) from the expert panelist compared to total MSK examination knowledge of the fellows	39 ± 16.2 (10–70)
Estimated percentage of time spent in direct patient care compared to total work hours	60 ± 12.2 (33–75)
Estimated percentage of patient encounters for shoulder pain compared to all patient counters	14 ± 6.1 (7–25)
Estimated percentage of use of the ultrasound of the shoulder for shoulder pain encounters	15 ± 29.2 (0–82)
Self-rating of the MSK examination skills from 0–100 (0 being no skills, 100 being best possible skills)	75 ± 10.6 (50–95)
Estimated percentage of shoulder examination skills learned during fellowship training compared to all shoulder examination learning experience	47 ± 20 (14–80)
1B. Qualitative background information, region (city, state)	
Fellowship training location where aforementioned shoulder examination skills were learned	Northeast: 7;South: 2;West: 1;Midwest: 1(Boston, Massachusetts; Worcester, Massachusetts; Farmington, Connecticut; New Haven, Connecticut; San Antonio, Texas; Dallas, Texas; Los Angeles, California; Ann Arbor, Michigan)
Postfellowship graduation location where majority of nonfellowship-based shoulder examination skills were learned	Northeast: 10;South: 1 (Burlington, Vermont; Lebanon, New Hampshire; Boston, Massachusetts; Worcester, Massachusetts; Providence, Rhode Island; Farmington, Connecticut; Pittsburgh, Pennsylvania; Winston-Salem, North Carolina)

* MSK, musculoskeletal.

Table 2. Shoulder examination maneuvers that met consensus threshold*

	Items met consensus threshold for tier 1	Items met consensus threshold for tier 2
Inspection	Adequate exposure (two shoulders without overlaying clothes)[a]; observation from front[a]; observation from behind[a]	N/A
Palpation	SC joint[a]; AC joint[a]; subacromial space[a]; acromion[b]; biceps (long)[b]; glenohumeral space[b]; clavicle[c]	Coracoid process[b]; spinous process of the C-spine[b]; scapular spine[c]
Active ROM	Abduction[a]; internal rotation[a]; external rotation[a]; forward flexion[a]	Neck flexion[b]; neck lateral rotation[b]; neck lateral flexion[b]; neck retroflexion or extension[b]; elbow pronation or supination[c]
Passive ROM	Abduction[a]; internal rotation[a]; external rotation[a]; forward flexion[a]	N/A
Resistance maneuvers	Resisted internal rotation[a]; resisted external rotation[a]; cross arm adduction test[a]; empty can (Jobe's) test[a]; painful arc test[a]; resisted abduction[b]; Speed's test[b]	Belly press test[a]; scapular winging test[a]; shoulder apprehension test[a]; Adson's test[a]; drop arm test[b]; Yergason's test[b]; Spurling's test[c]

* Consensus threshold definition: ≥70% (or ≥8/11 answers). Tier 1: teach all fellows and should be performed routinely. Tier 2: teach all fellows but should be performed only in specific scenarios. AC, acromioclavicular; N/A, not available; ROM, range of motion; SC, sternoclavicular.
[a] Met consensus during the first round.
[b] Met consensus during the second round.
[c] Met consensus during the third round.

anterior and posterior palpation. Panel met the consensus for anterior and posterior palpation to look for bulge sign, acknowledging that this is rarely evident except for lean patients with very large effusions.

Tier 2 palpation maneuvers. Palpation of the coracoid process was deemed a low-yield maneuver given the difficulty in palpating this structure in obese individuals and the unclear value if palpation causes pain. However, given its rare but potential involvement in tendonitis, it met the consensus threshold for tier 2. Palpation of the C-spine spinous processes met the consensus threshold for tier 2 because the expert panel felt that it should be done if tier 1 maneuvers fail to reveal the pathology. Lastly,

Table 3. Preferred methods for different shoulder examination maneuver techniques*

	Technique 1	Technique 2	Technique 3 (if available)
Palpation			
Biceps (long) palpation	With static humeral head (4/11)	With dynamic humeral head (7/11)	N/A
Glenohumeral space palpation[a]	Anterior/posterior palpation (9/11)	Anterior palpation only (2/11)	N/A
Scapular spine palpation[b]	Scapular spine only (8/11)	Scapular spine and scapular region (3/11)	N/A
Active ROM			
External rotation during active ROM[c]	With elbow at side (8/11)	With shoulder abducted at 90° (2/11)	By asking the patient to "reach behind the head" (1/11)
Passive ROM			
Abduction during passive ROM	Abduction to 90° (7/11)	Abduction to 180° (4/11)	N/A
Forward flexion during passive ROM[a]	Forward flexion to 30° (3/11)	Forward flexion to 180° (8/11)	N/A
Internal rotation during passive ROM	With elbow at side (0/11)	With shoulder abducted at 45–60° (5/11)	Shoulder abducted at 90° (6/11)
External rotation during passive ROM	With elbow at side (4/11)	With shoulder abducted at 45–60° (2/11)	Shoulder abducted at 90° (5/11)
Resistance maneuvers			
Empty can (Jobe's) test[a]	With shoulder forward extension to 30° (8/11)	With shoulder forward extension to 60° (3/11)	N/A
Hawkins Kennedy test	Both active and passive motion with shoulder forward flexion (3/11)	Just passive motion with shoulder abduction (1/11)	Just passive motion with shoulder forward flexion (7/11)
Neer's test[b]	Passive motion without arm pronation (1/11)	Passive motion with arm pronation (10/11)	Active motion without arm pronation (0/11)
Speed's test[a]	With simultaneous biceps palpation (3/11)	Without simultaneous biceps palpation (8/11)	N/A
Cross arm adduction test[b]	With passive motion (9/11)	With active motion (2/11)	N/A

* Consensus threshold definition: ≥70% (or ≥8/11 answers). N/A, not available; ROM, range of motion.
[a] Met consensus during the first round.
[b] Met consensus during the second round.
[c] Met consensus during the third round.

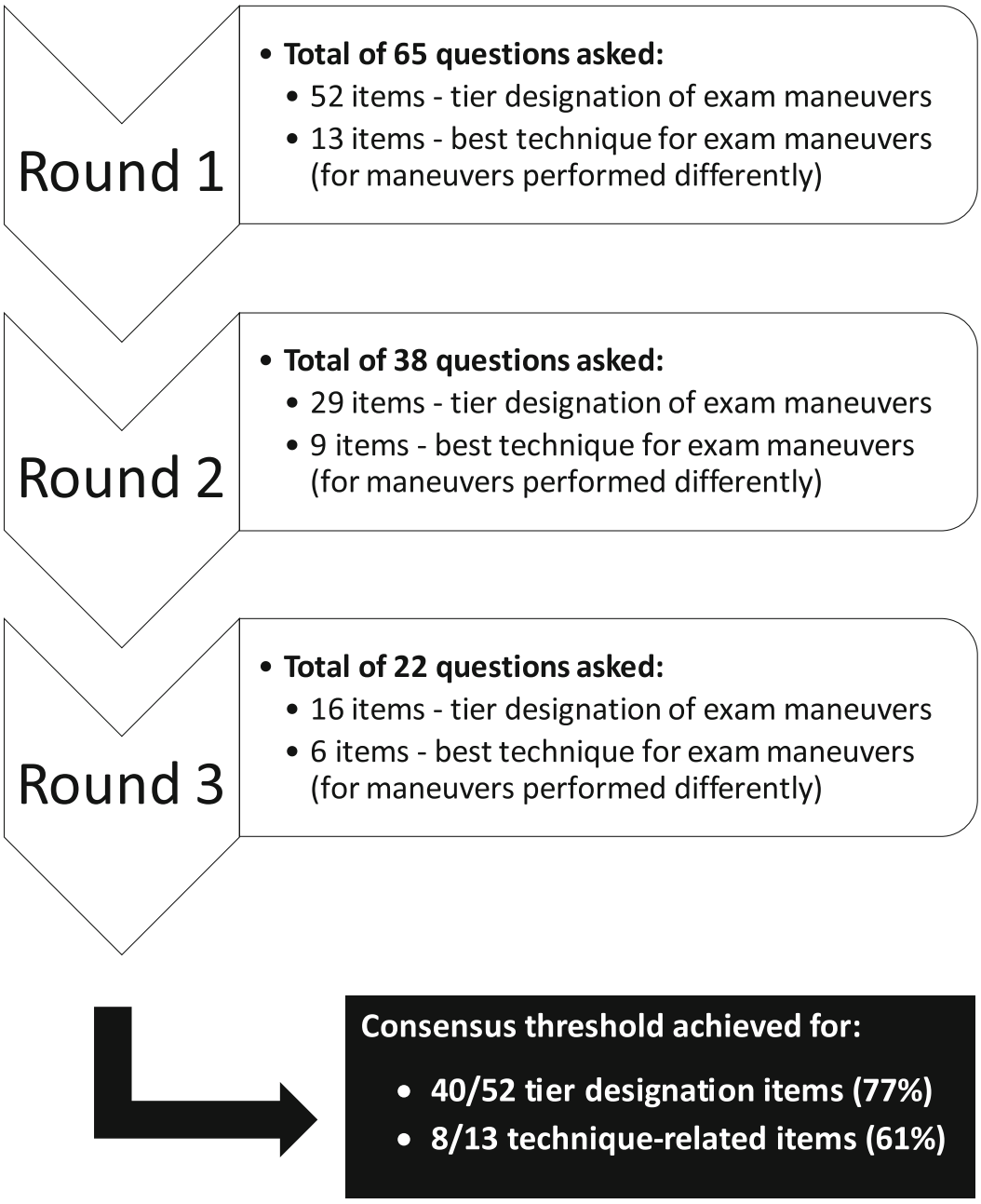


Figure 1. Study design: Delphi exercise.

palpation of the scapular spine was included in tier 2 only after the third round because it was deemed low yield for shoulder pain evaluation but useful as a landmark for glenohumeral joint injection and suprascapular nerve block.

Palpation maneuvers that did not meet the consensus threshold. Palpation of the deltoid muscle was noted as a useful maneuver for comparison of muscle bulk and evaluation of axillary neuropathy, but its overall value for the evaluation of isolated nonspecific shoulder pain was thought to be low yield. Palpation of the trapezius muscle, an essential structure to examine for suspected fibromyalgia or myofascial pain syndrome, did not meet the consensus threshold for a particular tier. Palpation of the elbow lateral epicondyle, with the idea of examining the joint adjacent to the painful joint, did not meet consensus for any tier.

Active ROM. *Tier 1 active ROM maneuvers.* All the listed maneuvers here met the consensus threshold during the first survey round, highlighting the yield of ensuring an adequate ROM. Trainees assessed the presence of a painful arc while asking their patients to abduct to 180°. The preferred method for internal and external rotation during active ROM testing included instructing

Table 4. Shoulder examination maneuvers that failed to meet consensus threshold*

	Examination maneuver
Inspection	N/A
Palpation	Deltoid muscle (tier 1: 7/11; tier 2: 3/11; tier 3: 1/11); trapezius muscle (tier 1: 6/11; tier 2: 5/11; tier 3: 0/11)[a]; elbow lateral epicondyle (tier 1: 1/11; tier 2: 6/11; tier 3: 4/11)
Active ROM	Retroflexion and extension (tier 1: 7/11; tier 2: 4/11; tier 3: 0/11)[a]
Passive ROM	Retroflexion and extension (tier 1: 6/11; tier 2: 4/11; tier 3: 1/11); elbow flexion and extension (tier 1: 2/11; tier 2: 6/11; tier 3: 3/11)
Resistance maneuvers	Hawkins Kennedy test (tier 1: 5/11; tier 2: 6/11; tier 3: 0/11)[a]; Neer's test (tier 1: 5/11; tier 2: 6/11; tier 3: 0/11)[a]; liftoff test (tier 1: 4/11; tier 2: 7/11; tier 3: 0/11)[a]; resisted elbow flexion (tier 1: 3/11; tier 2: 4/11; tier 3: 4/11); Hornblower's test (tier 1: 1/11; tier 2: 3/11; tier 3: 7/11); resisted neck flexion (tier 1: 1/11; tier 2: 3/11; tier 3: 7/11)

* Consensus threshold definition: ≥70% (or ≥8/11 answers).Tier 1: teach all fellows and should be performed routinely. Tier 2: teach all fellows but should be performed only in specific scenarios. Tier 3: teach only to selected fellows and should be performed only in specific scenarios. N/A, not available; ROM, range of motion.
[a] Maneuvers that did not receive any votes for tier 3, suggesting they should be taught to all fellows.

patients to keep elbows at their side while the elbows were flexed to 90°.

Tier 2 active ROM maneuvers. This group includes active ROM maneuvers of the neck with lateral rotation (turning head from side to side), lateral flexion (stretching to the shoulder with the ear), forward flexion (touching chest with their chin), extension (making their face parallel to the ceiling). Experts noted the high prevalence of cervical radiculopathy and necessity of checking neck ROM if tier 1 maneuvers are unrevealing, even in the scenario of isolated shoulder pain without weakness or numbness. Elbow pronation and supination, with the idea of examining the joint adjacent to the painful joint, met the consensus threshold for tier 2.

Active ROM maneuvers that did not meet consensus threshold. This group only included shoulder extension (retroflexion). Those who favored tier 1 suggested that it should be done routinely as it is part of the shoulder ROM as opposed to those who favored tier 2, emphasizing no added value because the main relevant pathologies of interest in rheumatology practice, glenohumeral arthritis, and adhesive capsulitis are better assessed with other ROM maneuvers.

Passive ROM. Expert panel emphasized that if a patient can perform a complete active ROM without discomfort, there is no need to conduct the passive tests. However, as this is rarely the case in real life, passive ROM maneuvers were deemed essential.

Tier 1 passive ROM maneuvers. All the maneuvers listed here met the consensus threshold during the first survey round, highlighting the yield of ensuring an adequate ROM. Notably, the cataloging of passive ROM maneuvers revealed significant variations in techniques among trainees.

Regarding forward flexion during passive ROM, the panel met the consensus threshold for favoring flexion to 180° over flexion to 30°. For abduction, the majority suggested checking until 90° of abduction over 180° of abduction, with an overarching theme that the 90° method isolates the glenohumeral joint better than the 180° of abduction technique, which requires additional scapulothoracic and acromioclavicular motion. For internal and external rotation, the panel highlighted the ease of noticing the limitation of ROM while the patient's shoulder is abducted (90° or 45° if full abduction is painful), compared to elbow at the side technique.

Passive ROM maneuvers that did not meet consensus threshold. Passive shoulder extension (retroflexion) had the following voting distribution: tier 1, 6 of 11; tier 2, 4 of 11; and tier 3, 1 of 11. People who favored tier 1 suggested that it should be done routinely because it is part of the shoulder ROM as opposed to people who favored tier 2, emphasizing no added value because the main relevant pathologies of interest in rheumatology practice, glenohumeral arthritis and adhesive capsulitis, are better assessed with other ROM maneuvers. Passive elbow flexion and extension, with the idea of examining the joint adjacent to the painful joint, failed to meet the consensus threshold for any tier.

Resistance and provocative maneuvers. *Tier 1 resistance and provocative maneuvers.* Following the first round, resisted internal rotation, resisted external rotation, empty can (Jobe's) test, cross arm adduction rest, and painful arc testing met the consensus threshold for tier 1. Following the second round, resisted abduction and Speed's testing were added to this list.

Among tier 1 provocative maneuvers, empty can, cross arm adduction, and Speed's testing were performed differently. Consensus was reached for the empty can maneuver to be performed with the shoulder at 30° forward from the coronal plane, to best isolate the supraspinatus tendon at the scapular plane. Cross arm adduction test was to be performed as a passive ROM rather than active ROM. Speed's testing was to be done with the elbow at full extension and no simultaneous biceps palpation, given the lack of added value with palpation.

Tier 2 resistance and provocative maneuvers. Following the first round, belly press test, scapular winging testing via wall push-up method, thoracic outlet testing via Adson's maneuver, and shoulder apprehension testing met the consensus threshold

for tier 2. After the second round, the drop arm test and Yergason's test were added to this list. Lastly, Spurling's test was added to the tier 2 after the third round.

Experts highlighted the benefit of a belly press test to confirm subscapularis pathology if initial resisted internal testing is positive or equivocal. Yergason's testing to evaluate biceps and/or superior labral anterior posterior (SLAP) pathologies deemed beneficial if initial Speed's testing is positive or equivocal. Spurling's maneuver was felt to be a useful supplement to neck active ROM maneuvers when evaluating for cervical radiculopathy.

Resistance and provocative maneuvers that did not meet consensus threshold. This list included liftoff testing for subscapularis pathologies, Neer's and Hawkins Kennedy maneuvers for subacromial impingement, and Hornblower's test for teres minor evaluation. Additionally, resisted neck forward flexion and resisted elbow flexion tests, failed to meet the consensus threshold for any tiers.

Final votes highlighted that the following maneuvers should be taught to all trainees: Hawkins Kennedy test (tier 1, 5 of 11; tier 2, 6 of 11), Neer's test (tier 1, 5 of 11; tier 2, 6 of 11), and liftoff test (tier 1, 4 of 11; tier 2, 7 of 11). For subacromial impingement testing via Hawkins Kennedy and Neer's tests, concerns were raised about their sensitivity (Sn) and specificity (Sp).

Additionally, Hawkins Kennedy and Neer's tests were performed by three different techniques for each of these maneuvers. For Hawkins Kennedy, the majority suggested performing the Hawkins Kennedy test as a passive motion when shoulder is at approximately 90° of forward flexion followed by forced internal rotation of the shoulder. For Neer's test, the consensus was met to do it as a passive motion with arm in pronation.

DISCUSSION

Physical examination evaluation of nonspecific isolated shoulder pain can be challenging and has significant variations among trainees and educators, likely due to the wide range of rheumatic and nonrheumatic pathologies that can affect the shoulder and the complex regional anatomy. This list of examination maneuvers, individual techniques, and their designation to different tiers provides high-yield information for rheumatology trainees and educators.

Our study has several strengths. First, all members of our expert panel are experienced rheumatologists and educators via their work as fellowship PDs. Second, all participants completed all survey rounds, with no dropout throughout the study period. Third, to our knowledge, this is the first study focusing on rheumatology trainees as the target audience for shoulder examination. Ranking of the different maneuver techniques via anonymized short video-clip–based survey question methodology strengthens our study because it minimizes the potential misunderstanding by experts on what is being voted on compared to if the techniques were just described in words and reduces bias on voting.

Our study did not evaluate all shoulder examination maneuvers used in clinical practice because of our desire to base guidance on what rheumatology educators actually teach. Hence, there are many other maneuvers described in the literature that are not assessed by this study. Furthermore, consensus by select experts does not equate to correctness or completeness. However, as evidence is lacking to derive a rheumatology-specific shoulder examination protocol, a consensus-based approach may be acceptable here.[8] The small size of our study group decreases the power of our findings, as differences between consensus for a maneuver or technique and lack of consensus could hinge on the difference of 7 versus 8 experts of the 11 supporting a given tier.

Sharing anonymous comments with experts between survey rounds may have influenced final tier designations of certain views, potentially leading to achievement of tier designations because it allowed experts to revisit their viewpoints and review counterarguments before their votes on unsolved participants. Furthermore, the use of the anonymized Delphi methodology gave each member of the expert panel equal authority. We acknowledge that shoulder examination might vary from academic practices to community practices. Our panel estimated that approximately 50% of the shoulder examination skills stem from fellowship training, and this always reflects the academic setting. Lastly, although we described the published literature on examination test characteristics cited by the panel, this study did not attempt to reevaluate measures of diagnostic accuracy for examination maneuvers or evaluate the appropriateness of further diagnostic studies.

The tier 1 examination maneuvers along with preferred techniques and their video demonstration can be used as a reference document for the routine evaluation of nonspecific shoulder pain in rheumatology practice. Maneuvers that met the consensus threshold for tier 2 as well as maneuvers that did not meet consensus threshold, yet for which the majority favored tier 1 or tier 2 include high-yield maneuvers that can be performed after tier 1 maneuvers.

Regarding maneuvers that received votes for tier 3 designation, these should be interpreted based on the background intention of creating a list for education to be most helpful for general rheumatology practice. Hence, the results might be different for training rheumatologists who will have a focus in sports medicine practice or those who will be working with a group of orthopedic surgeons (ie, "selected fellows"). Such practice focus–based variations exist not only in physical examination education but also in musculoskeletal imaging curricula.[9]

There are already excellent video demonstrations of shoulder examination, including but not limited to demonstrations by Stanford Medicine, FIFA, Oxford Medical Education, and SLC MSK Education.[10–14] In comparison, our video demonstrations have a

unique advantage to the rheumatology community, given their creation based on input from an expert panel of rheumatology PDs and their unique aim at rheumatology fellows.

These recommendations are most relevant to the rheumatology clinic setting and may not fully align with expected approaches to nondescript shoulder pain in other practices such as primary care, orthopedics, sports medicine, and physiatry. Experts note that even though most commonly encountered etiologies may overlap between different practices (eg, rheumatology and primary care),[4,5] direct extrapolation to other practices necessitate caution because subspecialists likely have additional subspecialty-focused differentials, which might affect their voting-based assessment of the role of different maneuvers. Hence, this consensus exercise is most relevant for the rheumatology community. A study from a community rheumatology practice based in the United Kingdom outlines the most common etiologies for shoulder pain as rotator cuff lesions (65%), pericapsular soft tissue pain (11%), acromioclavicular joint pain (10%), and referred pain from cervical spine (5%).[4]

We will review resistance maneuvers based on their target structure to share the panel's reasoning and literature cited by experts throughout this Delphi exercise (Table 5). Interestingly, the evaluated maneuvers assessing common causes of shoulder pain lack likelihood ratios above 10 and below 0.1 in most circumstances,[15] reflecting the imperfect nature of the commonly used physical examination maneuvers.

Acromioclavicular joint testing via cross arm adduction test (tier 1) has an Sn of 77% and an Sp of 79% for osteoarthritis.[16] The glenohumeral joint was assessed by tier 1 active and passive ROM maneuvers. Shoulder apprehension test met the consensus threshold for tier 2 during the first round. This test has an Sn range of 58% to 72% and an Sp of 96% for detecting glenohumeral instability and has an Sn range of 29% to 62% and an Sp range of 42% to 70% for detecting SLAP pathologies.[16] Respondents opposing tier 1 status emphasized the lower yield of checking for shoulder dislocation routinely at the rheumatology clinic setting.

Biceps and SLAP pathology resistance maneuvers included Speed's test (tier 1) and Yergason's test (tier 2). Speed's test has an Sn range of 49% to 71% and an Sp range of 60% to 85% to detect biceps pathologies and an Sn range of 28% to 60% with an Sp range of 38% to 76% to detect SLAP pathologies.[16] Yergason's test has an Sn range of 14% to 75% with an Sp range of 78% to 81% to detect biceps pathologies and an Sn range of 12% to 26% with an Sp range of 70% to 87% to detect SLAP pathologies.[16] Literature quoted by respondents reviewing the correlation of examination findings and ultrasonography findings highlighted the higher Sn of Speed's test but higher Sp of Yergason's test to detect biceps long-head pathologies, further supporting consensus designation of Speed's test as a tier 1 maneuver while assigning Yergason's test to the tier 2 list.[17]

Subacromial impingement maneuvers included checking for the painful arc during active abduction, resisted abduction, Hawkins Kennedy, and Neer's testing. Painful arc testing met the consensus for tier 1 during the first round, with literature reporting an Sn range of 49% to 75% and an Sp range of 33% to 80% to detect subacromial impingement regardless of the etiology.[16] The resisted abduction test met the consensus for tier 1 despite

Table 5. Pooled Sn, Sp, LR+, and LR– numbers for resistance maneuvers*

Tier designation	Maneuver	Target pathology	Sn (%)	Sp (%)	LR+	LR–	Reference
1	Cross arm adduction	AC joint osteoarthritis	77	79	3.67	0.29	Hegedus EJ, et al[16]
2	Shoulder apprehension	Glenohumeral instability	58–72	96	16.25	0.36	Hegedus EJ, et al[16]
2	Shoulder apprehension	SLAP pathologies	29–62	42–70	1.02	0.98	Hegedus EJ, et al[16]
1	Speed's test	Biceps (long) pathologies	49–71	60–85	2.14	0.56	Hegedus EJ, et al[16]
1	Speed's test	SLAP pathologies	28–60	38–76	0.98	1.00	Hegedus EJ, et al[16]
2	Yergason's test	Biceps (long) pathologies	14–75	78–81	2.10	0.71	Hegedus EJ, et al[16]
2	Yergason's test	SLAP pathologies	12–26	70–87	0.86	1.04	Hegedus EJ, et al[16]
1	Painful arc	Subacromial impingement	49–75	33–80	1.41	0.68	Hegedus EJ, et al[16]
1	Resisted abduction (pain)	Subacromial impingement	55	75	2.20	0.60	Hegedus EJ, et al[16]
1	Resisted abduction (weakness)	Subacromial impingement	38–58	20–50	0.74	1.49	Hegedus EJ, et al[16]
N/A	Hawkins Kennedy test	Subacromial impingement	58	67	1.76	0.63	Gismervik SØ, et al[18]
N/A	Neer's test	Subacromial impingement	59	60	1.48	0.68	Gismervik SØ, et al[18]
N/A	Hawkins Kennedy test and Neer's test both positive	Subacromial impingement	78	50	1.56	0.44	Hermans J, et al[19]
1	Empty can	Supraspinatus tear (full thickness)	76	39	1.25	0.62	Bak K, et al[21]
2	Drop arm	Supraspinatus tear (full thickness)	41	83	2.41	0.71	Bak K, et al[21]
1	Resisted external rotation	Infraspinatus tear or tendinopathy	46–84	53–100	2.71	0.46	Hegedus EJ, et al[16]
2	Belly press	Subscapularis	28–50	96–99	13	0.63	Hippensteel KJ, et al[22]
N/A	Liftoff	Subscapularis	12–25	95–100	6	0.85	Hippensteel KJ, et al[22]
2	Spurling's test	Cervical radiculopathy	30–52	74–96	2.73	0.69	Hippensteel KJ, et al[22]
2	Wall push-up	Scapulothoracic movement	89–100	0–8	0.98	1.50	Lohre R, et al[23]
2	Adson's test	Thoracic outlet syndrome	92	N/A	N/A	N/A	Sadeghi-Azandaryani M, et al[24]

* AC, acromioclavicular; LR–, negative likelihood ratio; LR+, positive likelihood ratio; N/A, not available; SLAP, superior labral anterior posterior; Sn, sensitivity; Sp, specificity.

the opposing respondents suggesting tier 2 designation, given its higher yield for deltoid muscle strength testing or deltoid enthesitis testing, relatively uncommon explanations for isolated shoulder pain. Hegedus et al[16] reported an Sn of 55% and an Sp of 75% for resisted abduction with pain and an Sn range of 38% to 58% and an Sp range of 20% to 50% for resisted abduction with weakness.

Hawkins Kennedy or Neer's tests did not meet the consensus threshold for tier 1 or tier 2, but it is important to note that none of the experts voted for tier 3 designation for these, reflecting the unified opinion that they should be taught to all fellows. In subacromial impingement, Gismervik et al[18] reported Hawkins Kennedy test with a pooled Sn of 58% and a pooled Sp of 67%, whereas Neer's test had a pooled Sn of 59% and a pooled Sp of 60%. Hermans et al[19] reported performances for combined positive testing for both Hawkins Kennedy and Neer's, with an Sn of 78% and an Sp of 50%.

We observed different examination techniques for Neer's and Hawkins Kennedy tests. For Neer's test, consensus was for a passive motion with arm in pronation, despite literature by respondents noting that the initial demonstration of this maneuver by Dr Neer did not include arm pronation.[20] Performing Hawkins Kennedy test as a passive motion while the shoulder is at approximately 90° of forward flexion followed by forced internal rotation of the shoulder was congruent with the description by Drs Hawkins and Kennedy.[20]

Rotator cuff testing maneuvers met tier 1 consensus included resisted internal rotation and external rotation and abduction and empty can. Maneuvers met consensus for tier 2 included belly press and drop arm tests. Liftoff test and Hornblower test did not meet consensus threshold for any tiers.

Supraspinatus maneuvers meeting consensus threshold included empty can test (tier 1) and a drop arm test (tier 2). Respondents felt drop arm was more appropriate for tier 2 because it is less sensitive (41%) than the empty can test (76%).[21] Infraspinatus was evaluated by the resisted external rotation (tier 1), for which Hegedus et al[16] report an Sn range of 46% to 84% and an Sp range of 53% to 100% for pain or weakness during resisted external rotation to detect infraspinatus tendinopathy or tears.

Subscapularis testing maneuvers included resisted internal rotation in tier 1, belly press testing in tier 2, and liftoff test that failed to meet the consensus threshold for any tiers. Respondents reported a low Sn but a high Sp of both belly press and liftoff testing, explaining why only resisted internal rotation met tier 1 consensus threshold. The belly press has a reported Sn range of 28% to 50%, an Sp range of 96% to 99%, positive likelihood ratio (LR+) range of 12.2 to 20.0, and the liftoff test has an Sn range of 12% to 25%, an Sp range of 95% to 100%, and an LR+ of 5.0 for subscapularis pathologies.[22] Teres minor testing via Hornblower's maneuver was deemed low yield because isolated teres minor pathology is rarely encountered in rheumatology practice.

Cervical radiculopathy evaluation with Spurling's testing, in addition to checking neck ROM, met the consensus threshold for tier 2, with literature reporting an Sn range of 30% to 52% and an Sp range[18] of 74% to 96%. Lastly, the wall push-up test, (scapular winging evaluation) with an Sn range of 89% to 100% to evaluate for serratus anterior muscle dysfunction and Adson's test with an Sn of 92% to evaluate for thoracic outlet syndrome were assigned[23,24] to tier 2.

This consensus study provides a list of examination maneuvers taught by participating rheumatology PDs in the New England area, along with a tier-based evaluation of several maneuvers, different examination techniques, and video demonstrations of key aspects for nonspecific shoulder pain evaluations in rheumatology practice. These findings are highly relevant to rheumatologists teaching the shoulder examination to rheumatology trainees.

ACKNOWLEDGMENTS

Mehmed Taha Dinc, an internal medicine resident at Boston University Medical Center at the time of the Supplementary video recording, agreed to volunteer to be a simulated patient for Supplementary videos. Benjamin Kissin, an undergraduate student at Boston University at the time of the Supplementary video editing, assisted with the editing and processing of the Supplementary videos.

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 Kissin 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. Dinnes J, Loveman E, McIntyre L, et al. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. Health Technol Assess 2003;7(29):1–166.
2. Bongers PM. The cost of shoulder pain at work. BMJ 2001; 322(7278):64–65.
3. Solomon DH, Bates DW, Schaffer JL, et al. Referrals for musculoskeletal disorders: patterns, predictors, and outcomes. J Rheumatol 2001;28(9):2090–2095.
4. Vecchio P, Kavanagh R, Hazleman BL, et al. Shoulder pain in a community-based rheumatology clinic. Br J Rheumatol 1995;34(5): 440–442.
5. Ostör AJ, Richards CA, Prevost AT, et al. Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology (Oxford) 2005;44(6):800–805.
6. Virtual Rheumatology Practicum. American College of Rheumatology. Accessed August 1, 2024. https://rheumatology.org/virtual-rheumatology-program
7. Rheumatology (IM) Fellowship Programs in New England. FREIDA AMA Residency and Fellowship Programs Database. 2024. Accessed

August 1, 2024. https://freida.ama-assn.org/search/list?spec=42821&loc=01

8. Minas H, Jorm AF. Where there is no evidence: use of expert consensus methods to fill the evidence gap in low-income countries and cultural minorities. Int J Ment Health Syst 2010;4(1):33.
9. Yinh J, Torralba KD, Choi KS, et al; for USSONAR. North American musculoskeletal ultrasound scanning protocol of the hip, knee, ankle, and foot: update of a Delphi consensus study. Clin Rheumatol 2021; 40(10):4233–4242.
10. How to Conduct a Shoulder Exam. Stanford Medicine 25. 2018. Accessed August 1, 2024. https://stanfordmedicine25.stanford.edu/the25/shoulder.html
11. FIFA. *Shoulder Examination | Practical clinical examination skills*. 2019. Accessed August 1, 2024. https://www.youtube.com/watch?v=MEopxiLeMtA
12. Oxford Medical Education. *Shoulder Examination – Orthopaedics*. 2012. Accessed August 1, 2024. https://www.youtube.com/watch?v=g8xtOqZFTwo
13. SLC MSK Education. Shoulder Physical Exam. 2015. Accessed August 1, 2024. https://www.youtube.com/watch?v=dzHFwkXBmvU&t=7s
14. Battistone MJ, Barker AM, Beck JP, et al. Validity evidence for two objective structured clinical examination stations to evaluate core skills of the shoulder and knee assessment. BMC Med Educ 2017; 17(1):13.
15. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004; 329(7458):168–169.
16. Hegedus EJ, Goode AP, Cook CE, et al. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med 2012;46(14):964–978.
17. Micheroli R, Kyburz D, Ciurea A, et al. Correlation of findings in clinical and high resolution ultrasonography examinations of the painful shoulder. J Ultrason 2015;15(60):29–44.
18. Gismervik SØ, Drogset JO, Granviken F, et al. Physical examination tests of the shoulder: a systematic review and meta-analysis of diagnostic test performance. BMC Musculoskelet Disord 2017;18(1):41.
19. Hermans J, Luime JJ, Meuffels DE, et al. Does this patient with shoulder pain have rotator cuff disease?: the Rational Clinical Examination systematic review. JAMA 2013;310(8):837–847.
20. Hawkins RJ, Kennedy JC. Impingement syndrome in athletes. Am J Sports Med 1980;8(3):151–158.
21. Bak K, Sørensen AK, Jørgensen U, et al. The value of clinical tests in acute full-thickness tears of the supraspinatus tendon: does a subacromial lidocaine injection help in the clinical diagnosis? A prospective study. Arthroscopy 2010;26(6):734–742.
22. Hippensteel KJ, Brophy R, Smith MV, et al. A comprehensive review of physical examination tests of the cervical spine, scapula, and rotator cuff. J Am Acad Orthop Surg 2019;27(11):385–394.
23. Lohre R, Elhassan B. Serratus anterior dysfunction examination: wall push-up or shoulder flexion resistance test? JSES Int 2022;6(5): 859–866.
24. Sadeghi-Azandaryani M, Bürklein D, Ozimek A, et al. Thoracic outlet syndrome: do we have clinical tests as predictors for the outcome after surgery? Eur J Med Res 2009;14(10):443–446.

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1273–1281
DOI 10.1002/acr.25558

American College of Rheumatology
Empowering Rheumatology Professionals

Recommendations for the Use of Disease-Modifying Antirheumatic Drugs in Pregnancy and Reproductive Health for Patients With Rheumatic Disease: A Scoping Review

Athena Chin,[1] Alice Terrett,[2] Mihye Kwon,[3] Samuel Whittle,[4] and Catherine Hill[5]

Objective. Autoimmune rheumatic diseases commonly affect individuals of childbearing age, with historically increased adverse pregnancy outcomes in this group. The advent of disease-modifying antirheumatic drugs (DMARDs) has fostered more suitable conditions for pregnancy; however, this is accompanied by challenges in ensuring safe use in reproductive health. The aim of this review is to compare existing guideline recommendations for the use of DMARDs in pregnancy and reproductive health for patients with rheumatic disease.

Methods. A scoping review was performed with Medline and Embase, in addition to a hand search, to identify guidelines published since 2014 by academic societies in rheumatology that addressed the management of DMARDs in pregnancy in any of rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, and systemic lupus erythematosus. Conventional synthetic DMARDs (csDMARDs) (methotrexate, sulfasalazine, leflunomide, and hydroxychloroquine), biologic DMARDs (bDMARDs) (adalimumab, etanercept, infliximab, golimumab, certolizumab, abatacept, tocilizumab, rituximab, and anakinra), and targeted synthetic DMARDs (tsDMARDs) (tofacitinib, baricitinib, and upadacitinib) were targeted. Two authors performed data extraction in duplicate (AC, AT).

Results. A total of 18 guidelines were included. Recommendations for DMARD use in preconception were present in 10 guidelines (56%), lactation in 12 guidelines (67%), and male fertility in 6 guidelines (33%). A total of 13 guidelines (72%) included recommendations for csDMARDs, 13 guidelines (72%) included recommendations for bDMARDs, and 5 guidelines (28%) included recommendations for tsDMARDs. There was moderate evidence supporting relatively uniform csDMARD recommendations, compared to minimal evidence for bDMARD and tsDMARD use with variable recommendations.

Conclusion. There is heterogeneity in the formulation of guidelines on the use of DMARDs in pregnancy. Recommendations for csDMARDs were similar between guidelines. There was significant variability in recommendations for bDMARD and tsDMARD use, reflecting current minimal literature in this area.

INTRODUCTION

Autoimmune rheumatic diseases (AIRDs) such as rheumatoid arthritis (RA), ankylosing spondylitis (AS), psoriatic arthritis (PsA), and systemic lupus erythematosus (SLE) commonly affect individuals of childbearing age and have historically had significant effects on reproductive health.[1] Patients with AIRDs traditionally have been observed to be at greater risk of adverse perinatal outcomes, including risk of preeclampsia, emergency Cesarean section, preterm delivery, and low birth weights.[1–3]

The advent of novel disease-modifying antirheumatic drugs (DMARDs) for treatment of these rheumatic conditions has significantly improved disease control and hence fostered more suitable conditions for pregnancy.[4] As such, there has been a rise over time in number of births among women with rheumatic

[1]Athena Chin, MBBS, FRACP: Royal Adelaide Hospital and University of Adelaide, Adelaide, South Australia, Australia, and Flinders Medical Centre, Bedford Park, South Australia, Australia; [2]Alice Terrett, MBBS: Royal Adelaide Hospital, Adelaide, South Australia, Australia, Flinders Medical Centre, Bedford Park, South Australia, Australia, and The Queen Elizabeth Hospital, Woodville South, South Australia, Australia; [3]Mihye Kwon, MBBS: Konyang University Hospital, Daejeon, South Korea; [4]Samuel Whittle, MBBS, MClinEpi, FRACP: The Queen Elizabeth Hospital, Woodville South, South Australia, Australia, and University of Adelaide, Adelaide, South Australia, Australia; [5]Catherine Hill, MBBS, MD, MSc, FRACP: Royal Adelaide Hospital and University of Adelaide, Adelaide, South Australia, Australia, and The Queen Elizabeth Hospital, Woodville South, South Australia, Australia.

Drs Whittle and Hill are co-senior authors and contributed equally to this work.

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25558.

Address correspondence via email to Athena Chin, MBBS, FRACP, at athena.chin@sa.gov.au.

Submitted for publication December 13, 2024; accepted in revised form April 11, 2025.

SIGNIFICANCE & INNOVATIONS

- There is relative uniformity in guidelines regarding the use of conventional synthetic disease-modifying antirheumatic drugs (DMARDs) in pregnancy in patients with rheumatic diseases.
- There has been a gradual shift toward recommendations of certain biologic DMARD use in reproductive health in the last 10 years.
- There is a lack of guidelines regarding DMARD use in male fertility and newborn health.
- Greater consumer representation is important in future formulation of guidelines in this area.

disease, and, notably, risks of adverse pregnancy and low fetal birth weight have reduced.[4,5]

However, this is accompanied by new challenges in ensuring safe use in reproductive health, with concerns regarding teratogenicity, safety in breastfeeding, and the preconception state. Certain DMARDs have clear evidence of potential harm in pregnancy; leflunomide has been observed to be embryotoxic in preclinical animal studies,[6] and methotrexate is a current standard medical treatment for ectopic pregnancy.[7] There are limited data regarding the use of newer agents such as biologic DMARDs (bDMARDs) and targeted synthetic DMARDs (tsDMARDs) in reproductive health, although there is emerging literature to support the use of bDMARDs for disease control in women with inflammatory arthritis during pregnancy.[8] Similarly, given DMARDs are immunosuppressive by nature of mechanism of disease control, implications for newborn immunity requires consideration.[9]

The potential for unknown teratogenic or developmental implications in children is always challenging in the use of novel drugs during pregnancy. However, the use of DMARDs in reproductive health must be balanced with the benefits in achievement of parental disease control, which appears to cultivate positive peripartum outcomes.[10,11] This review aims to compare existing clinical practice guideline recommendations for the use of DMARDs in pregnancy and reproductive health for patients with rheumatic disease.

MATERIALS AND METHODS

Type of review and search strategy. We sought to identify gaps in literature in this area by reviewing a broad range of sources, and therefore a scoping review was performed following established guidelines from the Joanna Briggs Institute Scoping Review Methodology Group.[12] Two main searches for relevant publications from Medline and Embase databases were performed, with support from research librarians. A single reviewer (MK) performed the initial title and abstract prescreening, with subsequent further screening by two reviewers (AC and AT). Additional guidelines were identified from hand search of reference lists of guidelines included. Full-text reviews were then performed by two reviewers (AC and AT), and articles were assessed against predetermined inclusion and exclusion criteria. Inclusion criteria were the following: (1) recommendations, guidelines type, and published within the last 10 years (2014–2024); (2) author or group of national or international academic societies in rheumatology; (3) target disease and patients (adult patients with RA, AS, PsA, juvenile idiopathic arthritis, or SLE); and (4) target drugs including conventional synthetic DMARDs (csDMARDs) (methotrexate, sulfasalazine, leflunomide, hydroxychloroquine, and minocycline), bDMARDs (adalimumab, etanercept, infliximab, golimumab, certolizumab, abatacept, tocilizumab, rituximab, sarilumab, anakinra, and secukinumab), and tsDMARDs (tofacitinib, baricitinib, upadacitinib, and filgotinib). The exclusion criteria were the following: (1) systematic reviews and scoping reviews, meta-analyses, observational studies, and randomized controlled trials; (2) author or group of individual hospitals or research groups; (3) target drugs (publications exclusively focused on immunosuppressants); and (4) no English translation available. Any discrepancies were resolved by discussion.

We searched Medline and Embase, in addition to a hand search of references, to identify the most recent versions of guidelines by national or international academic societies in rheumatology that addressed management of DMARDs in pregnancy in any of adult patients with RA, AS, PsA, and SLE. csDMARDs (methotrexate, sulfasalazine, leflunomide, and hydroxychloroquine), bDMARDs (adalimumab, etanercept, infliximab, golimumab, certolizumab, abatacept, tocilizumab, rituximab, anakinra, and secukinumab), and tsDMARDs (tofacitinib, baricitinib, and upadacitinib) were targeted. Guidelines published since 2014 were included.

Data extraction. Two authors (AC and AT) performed data extraction in duplicate. Data were extracted based on pre-established characteristics of interest, including year of publication, country or region of publication, included rheumatic conditions, included DMARDs, inclusion of guidelines on specific areas of reproductive health (breastfeeding, preconception, male fertility, and newborn health), conflict of interest declaration and management, panel structure, method of formulation of recommendation, and evidence system used. Guideline recommendations regarding DMARD use in reproductive health were extracted, with exclusion of in-text recommendations. Review and discussion between the two authors following initial extraction was performed to further refine data.

RESULTS

Literature search. A total of 743 unique articles were identified from the search strategy. A hand search identified a further two articles, and full-text review was performed on a total of

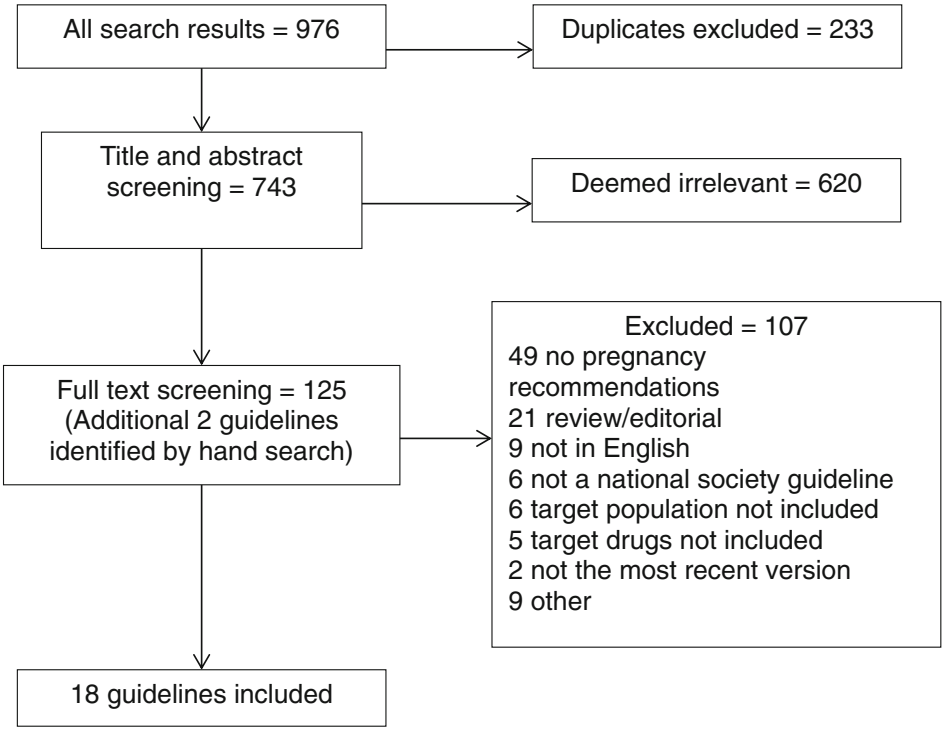


Figure 1. Flowchart of included studies.

125 articles following title and abstract screening. A total of 107 articles were excluded after full-text review for the following reasons: no pregnancy recommendations (n = 49); review, editorial, or summary article (n = 21); not available in English (n = 9); not an international or national society guideline (n = 6); target population not included (n = 6); target drug not included (n = 5); not the most recent version of society guideline (n = 2); and other (including article subsequently retracted, comparison of guidelines, and defining scope for future guidelines; n = 9). The final review included 18 articles (Figure 1).

Characteristics of included guidelines. The 18 guidelines were published by national and international societies from a range of regions, including 7 guidelines from Europe (38.9%),[13–19] 5 guidelines from Asia (27.8%),[20–24] 4 guidelines from North America (22.2%),[25–28] 1 guideline from South America (5.6%),[29] and 1 guidelines from Africa (5.6%).[30] In addition to pregnancy, recommendations on DMARD use in breastfeeding were included in 13 guidelines (72.2%),[13–18,20–22,25–28] preconception in 10 guidelines (55.6%),[13–15,17,20,22,25–28] male fertility in 6 guidelines (33.3%),[14–17,25,28] and postpartum newborn health in 2 guidelines (11.1%).[14,27] Four guidelines (22.2%) were a standalone guideline for the use of DMARDs in pregnancy.[13,14,25,28] More than half of the guidelines (n = 10; 55.6%) were RA-specific guidelines,[15–17,20–24,26,27] and two guidelines (11.1%) were specific to SLE.[19,29] Most guidelines addressed csDMARDs[13–15,17,18,20,22,25–30] (n = 13; 72.2%) and/or bDMARDs[13,14,16,19–21,23–28,30] (72.2%), and only six guidelines (33.3%)[13,14,16,20,21,27] had recommendations for tsDMARDs.

Formulation of recommendations. The identified guidelines had a broadly homogenous approach to guideline formulation, with most performing a literature review before using panel voting to reach consensus for recommendations. All guidelines that specified a predefined level of agreement[13,14,17,19,21,23,24,30] (n = 8; 44.4%) required panel consensus >75% for a recommendation to be published. The Delphi or modified Delphi method[31] was used in five guidelines (27.8%).[13,15,24,27,30] All voting panels included rheumatologists. A range of other professions were also included, including medical specialists (obstetricians, geneticists, internal medicine physicians, ophthalmologists, epidemiologists, pediatric neonatologists, infectious diseases specialists, nephrologists, dermatologists, and cardiologists), allied health professionals, and research fellows. Consumers were involved in the development of five guidelines[13,14,25,26,30] (27.8%). Funding was disclosed by eight[13,15,20,21,23,25,28,29] societies (44.4%), and three[20,28,29] societies (16.7%) received pharmaceutical funding. The Appraisal of Guidelines for Research & Evaluation II[32] tool was used by three societies[21,23,28] (16.7%) during the formulation of recommendations.

Rating of evidence. There was a varied approach to rating of evidence within these guidelines. The most used evidence rating system was Grades of Recommendations, Assessment,

Development and Evaluation (GRADE)[33] (n = 7; 38.9%), then the Oxford Centre for Evidence-Based Medicine Levels of Evidence[34] (n = 2; 11.1%). Other systems used were the classification system by Shekelle et al,[35] Harbour and Miller,[36] and Scottish Intercollegiate Guidelines Network Grading System[37] (n = 1 each; 5.6%). Two societies (11.1%) based their recommendations on previously published guidelines, and four societies (22.2%) did not specify rating of evidence.

Recommendations in pregnancy. *Recommendations for csDMARDs.* A total of 13 guidelines[13–15,17,18,20,22,25–30] included recommendations for csDMARD use in pregnancy (Table 1). There was 100% agreement among these recommendations between guidelines, with a moderate level of evidence referenced. All guidelines recommended that methotrexate and leflunomide were contraindicated in pregnancy. Hydroxychloroquine and sulfasalazine were considered safe in pregnancy, with recommendation to continue. Two guidelines[20,28] (11.1%) specifically recommended supplemental folate supplementation with sulfasalazine use in pregnancy.

Recommendations for bDMARDs. Thirteen guidelines[13,14,16,19–21,23–28,30] included recommendations for bDMARD use in pregnancy. There was a low level of evidence for these recommendations. Recommendations differed between guidelines, however, there was a high level of agreement within individual panels. Three guidelines[19,21,28] published before 2018 recommended avoiding all bDMARDs in pregnancy due to lack of evidence, without recommendations for specific medications. The other 10 guidelines all recommended continuing tumor necrosis factor (TNF) inhibitors (adalimumab, golimumab, etanercept, infliximab, and certolizumab); 7 guidelines were conditional recommendations.[13,16,23,25–27,30] Certolizumab was strongly recommended to be continued throughout the entire pregnancy by five guidelines,[14,16,20,24,25] and three guidelines specified continuing certain TNF inhibitors (adalimumab, golimumab, and etanercept) for the first two trimesters only.[20,24,26]

Most guidelines recommended cessation in pregnancy due to lack of evidence when specifically discussing other biologic agents, including abatacept (n = 4),[13,16,20,25] tocilizumab (n = 5),[13,16,20,25,27] and anakinra (n = 3)[13,20,25]; however, one guideline had a conditional recommendation for continuing these medications in pregnancy "if no other drug [is] available,"[14] and another recommended they "should be used during pregnancy only when no other pregnancy-compatible drug can effectively control maternal disease."[13] Two guidelines included recommendations on secukinumab use in pregnancy[14,25]; one guideline recommended discontinuation,[25] and one recommended conditional use if no other medication was suitable.[14] Rituximab was specifically addressed in six guidelines, with half (n = 3)[13,16,20] recommending cessation and the others (n = 3)[14,25,27] conditionally recommending continuing if no alternate option and required for disease control in pregnancy.

Recommendations for tsDMARDs. Six guidelines[13,14,16,20,21,27] included recommendations for tsDMARD use in pregnancy. There was low-level to no evidence in this area, and all guidelines recommended avoiding tsDMARDs due to insufficient data on use in pregnancy. These recommendations had a high level of agreement within individual panels.

Recommendations for the preconception state. *Recommendations for csDMARDs.* Ten guidelines[13–15,17,20,22,25–28] included preconception recommendations for csDMARD use. Evidence was limited to cohort studies, with a high level of agreement within individual panels. These guidelines all had strong recommendations regarding methotrexate, with eight guidelines (80%) recommending cessation three months before conception[13,15,17,25–28] and two guidelines (20%) recommending cessation at least one month before.[14,22]

Six guidelines included recommendations for leflunomide,[13,14,20,25,27,28] and all recommended either a washout with cholestyramine or cessation two years before conception, or both. A blood test to check drug levels before conception was recommended by three of these guidelines (50%).[25,27,28] Two guidelines included recommendations for preconception use of sulfasalazine and hydroxychloroquine[14,25]; continuation of these drugs was recommended by both societies.

Recommendations for bDMARDs. Four guidelines included preconception recommendations for bDMARD use[13,20,25,26]; notably, one based their recommendations on another included guidelines' recommendations.[20] There was low-level evidence to support these recommendations. Regarding TNF inhibitors, three guidelines had conditional recommendations to continue these drugs preconception,[13,25,26] with two guidelines sharing a strong recommendation to continue certolizumab specifically. All four guidelines did not recommend continuing rituximab before conception, one guideline specified cessation 12 months before,[26] and one guideline specified 6 months before.[20] Two guidelines recommended discontinuing abatacept before conception,[13,26] and one guideline specified 10 weeks before conception.[26] The same two guidelines recommended ceasing tocilizumab; one guideline specified three months (5.5 half-lives) before conception.[26] One recommended cessation of anakinra before conception.[13] Notably, one guideline conditionally recommended continuing abatacept, tocilizumab, anakinra, and secukinumab during the preconception stage and ceasing once pregnant.[25]

Recommendations for tsDMARDs. Two guidelines included preconception recommendations for tsDMARD use, specifically tofacitinib. Both recommended cessation in the preconception stage, with differing time frame (two months before[13] vs two weeks before[14]).

Recommendations for breastfeeding. *Recommendations for csDMARDs.* Eleven guidelines included recommendations for csDMARD use in breastfeeding,[13–15,17,18,20,22,25–28]

Table 1. Recommendations for disease-modifying antirheumatic drug use in pregnancy*. Color table can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25537/abstract

	SSZ	HCQ	MTX	LEF	TNFi*	CTZ	ABA	TCZ	RTX	ANA	SEC
Saavendra Salinas, 2015 Mexico	●	●	●	●	●	●	●	●	●		
Skorpen, 2016 Europe [EULAR]	●	●	●	●	●	●	●	●	●	●	
Louthrenoo, 2016 Thailand					●	●	●	●	●		
Alpizar-Campos, 2016 Costa Rica	●	●	●	●	●	●		●	●		
Garcia-Vicuna, 2017 Spain			●								
Duarte, 2017 Portugal			●								
Kleinmann, 2017 France					●	●	●	●	●		
Lau, 2018 Hong Kong					●	●					
Pons-Estal, 2018 South America		●									
Kameda, 2019 Japan			●								
Ho, 2019 Hong Kong					#	●					
Alhajeri, 2019 Kuwai	●	●	●	●	#	●	●	●	●	●	
Sammaritano, 2020 America	●	●	●	●	●	●	●	●	●	●	●
Cardiel, 2021 Mexico	●	●	●	●	#	●					
Fiehn, 2021 Germany		●									
Fernandes, 2022 Portugal					●	●	●	●	●		
Russell, 2023 United Kingdom	●	●	●	●	●	●	●	●	●	●	●
Jaouad, 2023 Morocco	●				●	●					

MTX = methotrexate; SSZ = sulfasalazine; HCQ = hydroxychloroquine; LEF = leflunomide; TNFi = tumour-necrosis factor inhibitors; CTZ = certolizumab; ABA = abatacept; TCZ = tocilizumab; RTX = rituximab; ANA = anakinra; SEC = secukinumab

*TNFinhibitors that are not certolizumab (adalimumab, etanercept, golimumab, infliximab)
Yes for 1st + 2nd trimester

● Recommend use
● Conditional recommendation
● Recommend against use

with low-to-moderate levels of evidence available (Table 2). There were uniform recommendations for methotrexate (n = 10)[13–15,17,20,22,25–28] and leflunomide (n = 7)[13,14,20,25–28] that these two medications were not safe for use in lactation. Guidelines also had shared recommendations to continue hydroxychloroquine (n = 7).[13,14,18,20,25–27] Although there was also consensus that sulfasalazine was compatible in breastfeeding (n = 5)[13,14,20,25,26] for healthy, full-term infants, one of these guidelines specifically advised precaution in use with infants who had glucose-6-phosphate dehydrogenase deficiency, hyperbilirubinemia, or prematurity.[26]

Recommendations for bDMARDs. Eight guidelines included recommendations for bDMARD use in breastfeeding.[13,14,16,20,21,25,27,28] Three guidelines, all published before 2017, recommended globally avoiding bDMARDs due to insufficient evidence.[21,27,28]

The remaining five guidelines uniformly recommended that TNF inhibitor use was appropriate during breastfeeding, however, there were different recommendations for use of other bDMARDs during lactation. Notably, one article[13] stated that there was high disagreement between experts (25–31%) on clinical use of bDMARDs without data on transfer into breastmilk (abatacept

Table 2. Recommendations for disease-modifying antirheumatic drug use in breastfeeding*. Color table can be viewed in the online issue, which is available at http://onlinelibrary.wiley.com/doi/10.1002/acr.25537/abstract

	SSZ	HCQ	MTX	LEF	TNFi*	CTZ	ABA	TCZ	RTX	ANA	SEC
Saavendra Salinas, 2015 Mexico			🔴	🔴	🔴	🔴	🔴	🔴	🔴		
Skorpen, 2016 Europe [EULAR]	🟢	🟢	🔴	🔴	🟢	🟢	🔴	🔴	🔴	🔴	
Louthrenoo, 2016 Thailand					🔴	🔴	🔴	🔴	🔴	🔴	
Alpizar-Campos, 2016 Costa Rica		🟢	🔴	🔴	🔴	🔴	🔴	🔴	🔴	🔴	
Garcia-Vicuna, 2017 Spain			🔴								
Duarte, 2017 Portugal			🔴								
Kameda, 2019 Japan			🔴								
Alhajeri, 2019 Kuwai	🟢	🟢	🔴	🔴	🟢	🟢	🔴	🔴	🔴		
Sammaritano, 2020 America	🟢	🟢	🔴	🔴	🟢	🟢	🟡	🟡	🟢	🟡	
Cardiel, 2021 Mexico	🟢	🟢	🔴	🔴							
Fiehn, 2021 Germany		🟢									
Fernandes, 2022 Portugal					🟢	🟢	🔴	🔴	🔴		
Russell, 2023 United Kingdom	🟢	🟢	🔴	🔴	🟢	🟢	🟡	🟡	🟡	🟡	🟡

MTX = methotrexate; SSZ = sulfasalazine; HCQ = hydroxychloroquine; LEF = leflunomide; TNFi = tumour-necrosis factor inhibitors; CTZ = certolizumab; ABA = abatacept; TCZ = tocilizumab; RTX = rituximab; ANA = anakinra; SEC = secukinumab

*TNFinhibitors that are not certolizumab (adalimumab, etanercept, golimumab, infliximab)

🟢 Recommend use
🟡 Conditional recommendation
🔴 Recommend against use

and tocilizumab). Use of abatacept and tocilizumab in breastfeeding was conditionally recommended in two guidelines,[14,25] and secukinumab was felt compatible with breastfeeding in one guideline.[14] Anakinra was also conditionally recommended in two guidelines[14,25]; one guideline recommended avoidance due to no data.[13] Rituximab had varying recommendations among guidelines; one guideline strongly recommended this was compatible,[25] one guideline provided a conditional recommendation to continue,[14] and three guidelines specifically advised avoidance in lactation.[13,16,20]

Recommendations for tsDMARDs. Five guidelines included recommendations for tsDMARD use in breastfeeding.[13,14,20,21,27] [13,14,20,21,27] There are no current data available, and all guidelines recommended avoidance during lactation, with a high level of agreement among panels.

Recommendations for male fertility. *Recommendations for csDMARDs.* Five guidelines included recommendations for csDMARD use in male fertility.[14,15,17,25,28] Four guidelines included conditional recommendations for methotrexate; three guidelines (75%) recommended continuation of methotrexate with a moderate level of evidence,[14,17,25] with one guideline specifying "paternal exposure to low-dose [at <25 mg/wk] [methotrexate] is compatible with pregnancy."[14] One guideline recommended cessation of methotrexate three months before conception, with a low-grade level of evidence (based on expert opinion).[15] Leflunomide also had differing recommendations, with two guidelines recommending conditional continuation in men (2 of 3, 66.7%),[14,25] and one recommending discontinuation two years before or washout with cholestyramine and measurement of drug levels before conception in "both parents (men and women)…planning pregnancy."[28] In two guidelines addressing hydroxychloroquine and sulfasalazine use in male fertility,[14,25] there was shared strong recommendation for continuing hydroxychloroquine and conditional continuation of sulfasalazine. However, one guideline suggested performing semen analysis and ceasing sulfasalazine if conception is delayed by 12 months.[25]

Recommendations for bDMARDs. Three guidelines included recommendations for bDMARD use in male fertility.[14,16,25] There was a low level of evidence available for these recommendations.

All three guidelines recommended conditionally continuing rituximab in men. When included, there was also 100% agreement of recommendations to continue TNF inhibitors (2 of 2), anakinra (2 of 2), abatacept (2 of 2), and tocilizumab (2 of 2) in men trying to conceive.

Recommendations for tsDMARDs. Two guidelines included recommendations for tsDMARD, specifically tofacitinib, use in male fertility.[14,16] One guideline recommended continuation, although with limited evidence,[14] and one guideline recommended avoidance due to insufficient data.[16]

Recommendations for postpartum newborn health. Two guidelines[14,27] included recommendations for newborn health, specifically pertaining to maternal bDMARD use. Both had shared recommendations to discontinue TNF inhibitors (except certolizumab) in the third trimester to avoid risk of infection in the newborn; one guideline suggested ceasing all at 30 weeks,[27] and the other had more specific recommendations[14] (adalimumab cessation at 28 weeks gestation, etanercept cessation at 32 weeks gestation, golimumab cessation at 28 weeks gestation, and infliximab cessation at 20 weeks gestation). One guideline had recommendations regarding newborn vaccinations[14]; if abatacept, tocilizumab, rituximab, anakinra, or TNF inhibitors (except certolizumab) are used in the third trimester, all live vaccinations should be avoided until the infant is six months old. The same guideline recommended that certolizumab may be safely used throughout pregnancy without alteration to infant vaccination schedule. There were no specific recommendations pertaining to maternal use of csDMARDs or tsDMARDs and newborn health.

Other recommendations for drugs outside of inclusion criteria. *Glucocorticoids.* Seven guidelines[13,14,25–29] had recommendations regarding the use of glucocorticoids in pregnancy. There was a moderate level of evidence, with six guidelines providing conditional recommendations for use of glucocorticoids such as prednisolone in pregnancy. All of these guidelines recommended using the lowest dose possible, and two guidelines specifically recommended daily dose equivalent of <20 mg of prednisolone.[14,25] One guideline specific to SLE and antiphospholipid syndrome recommended avoiding routine use of glucocorticoids in pregnancy.[29]

Limitations. This scoping review was limited by the available low-quality evidence and minimal existing literature of use of DMARDs in reproductive health, in particular, for newer medications. The lack of evidence restricted recommendation formulations for bDMARDs and tsDMARDs, with generation of mostly conditional recommendations only. There was also distinct paucity of recommendations for DMARD use in male fertility and newborn health. Our review also only included guidelines from rheumatologic societies and hence broadening the scope to include other specialties that manage autoimmune conditions (such as renal and gastroenterology) would strengthen a future review of existing data and recommendations for DMARD use in reproductive health.

DISCUSSION

Guidelines for the use of DMARDs in reproductive health are readily available, however, they vary regarding specific recommendations. Evidence for use of drugs in pregnancy is often scant, given significant ethical implications for controlled trials in this cohort. Consideration is required for both maternal and fetal outcomes, and many guidelines rely on preclinical research, cohort studies, and retrospective review of drug exposure during pregnancy to formulate recommendations.

This review identified that there were more uniform recommendations for older medications (csDMARDs) between guidelines, likely due to greater existing literature of use in reproductive health. Recommendations regarding bDMARDs varied significantly between guidelines, with a clear temporal trend observed to favor more conditional recommendations for use in newer guidelines. This is observed over a 10-year period (2014–2024) and reflects growing literature available from retrospective cohort studies of use and exposure in peripregnancy settings, in addition to longevity in market of medications. The earliest class of bDMARDs approved by the US Food and Drug Administration were TNF inhibitors (etanercept; 1998), and guidelines from 2019 appear to favor use in pregnancy, compared to more recently approved bDMARDs such as abatacept (2005), which guidelines up to 2023 still recommended avoiding in the peripartum setting. Given biologic drugs are recombinant proteins typically similar to maternal IgG antibodies, there is varying ability for these drugs to physically cross the placenta and have direct effect on the fetus,[8] and emerging research on placental transfer may also be contributing to the gradually increasing confidence in use of DMARDs when necessary for disease control. A preliminary update of EULAR Points to Consider for Use of Antirheumatic Drugs in Reproduction, Pregnancy, and Lactation has been recently presented[38] and includes recommendations in favor of the use of all TNF inhibitors in pregnancy and the use of other bDMARDs such as abatacept, rituximab, tocilizumab, anakinra, and secukinumab if needed to control maternal disease. These significant changes since the previous points to consider published in 2016 further support the observation that relatively rapid changes in recommendations are occurring in this field.

Currently, the newest agents, tsDMARDs, had very limited recommendations due to a lack of evidence. Notably, tsDMARDs are used in other autoimmune conditions such as inflammatory bowel disease (IBD), and there is emerging registry data on tsDMARD exposure during pregnancy in this group of patients,[39] with recent Pregnancy in IBD and Neonatal Outcomes

consensus statements favoring the use of JAK inhibitors if "no viable alternatives to maintain maternal health."[40]

Although there was overall uniformity among drug recommendations in the preconception state, guidelines differed in duration that a medication needed to be held before pregnancy. Whereas timing of drug cessation before conception was specified, only one guideline clearly stated this was in relation to drug half-life.[26] There were scant recommendations pertaining to DMARD use in male fertility, and, interestingly, there was general agreement among guidelines on bDMARD use but different recommendations for csDMARD use in this setting, with overall very limited data available. The preliminary updated EULAR points to consider provides a broader list of DMARDs that have not demonstrated a clinically relevant impact on fertility or pregnancy outcome,[38] including sulfasalazine and methotrexate, highlighting the value of ongoing research on the impact of these drugs on sperm.[41]

Only two guidelines discussed newborn health guidelines with maternal bDMARD use,[14,27] and only one guideline provided recommendations on alteration of vaccination schedules in newborns.[14] Although it is necessary to exercise caution in use of any new medication in the context of pregnancy and potential implications on fetal developmental outcomes, initial studies suggest that there is no interference of bDMARD use up to infancy[42]; however, more literature is required to strengthen recommendations in this area.

The process of formulating recommendations was similar among guidelines; however, only five guidelines included consumers on their voting panel. Notably, women with autoimmune disease are particularly vulnerable during pregnancy, with high rates of ambivalence and depression,[43] and a retrospective study in women with AIRD revealed several unmet needs in knowledge about reproductive issues, with concerns that both active disease and medications required to control disease may harm the newborn.[44] This is supported by a study by Birru Talabi et al[45] that found up to 80% of women chose to discontinue DMARDs, including nonfetotoxic agents, and risked enduring inflammatory arthritis due to concerns of safety during pregnancy, despite emerging literature that there is minimal safety concerns with use of bDMARDs in pregnancy but higher risk of adverse pregnancy outcomes with uncontrolled inflammatory disease.[8] This highlights the need for consumer representation and integration of the patient perspective in formulation of these guidelines to support personalized discussions between clinicians and patients, as counseling and a multidisciplinary approach appears to herald a positive effect on family planning and confidence in decision-making.[44,46]

ACKNOWLEDGMENTS

Open access publishing facilitated by Flinders University, as part of the Wiley - Flinders University agreement via the Council of Australian University Librarians.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Chin 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. Kerola AM, Palomäki A, Laivuori H, et al. Patterns of reproductive health in inflammatory rheumatic diseases and other immune-mediated diseases: a nationwide registry study. Rheumatology Oxford 2024;63(10):2701–2710.
2. Sugawara E, Kato M, Fujieda Y, et al. Pregnancy outcomes in women with rheumatic diseases: a real-world observational study in Japan. Lupus 2019;28(12):1407–1416.
3. Strouse J, Donovan BM, Fatima M, et al. Impact of autoimmune rheumatic diseases on birth outcomes: a population-based study. RMD Open 2019;5(1):e000878.
4. Wallenius M, Salvesen KÅ, Daltveit AK, et al. Reproductive trends in females with inflammatory joint disease. BMC Pregnancy Childbirth 2016;16(1):123.
5. Wallenius M, Salvesen KÅ, Daltveit AK, et al. Secular trends of pregnancies in women with inflammatory connective tissue disease. Acta Obstet Gynecol Scand 2015;94(11):1195–1202.
6. Chambers CD, Johnson DL, Robinson LK, et al; Organization of Teratology Information Specialists Collaborative Research Group. Birth outcomes in women who have taken leflunomide during pregnancy. Arthritis Rheum 2010;62(5):1494–1503.
7. Verberne EA, de Haan E, van Tintelen JP, et al. Fetal methotrexate syndrome: a systematic review of case reports. Reprod Toxicol 2019;87:125–139.
8. Giles I, Thorne I, Schmidt NS, et al. The time of equipoise on the use of biological DMARDs in for inflammatory arthritis during pregnancy is finally over: a reappraisal of evidence to optimise pregnancy management. Lancet Rheumatol 2024;6(8):e546–e559.
9. Goulden B, Chua N, Parker E, et al. P202 administering live vaccines to infants exposed to biologic and targeted synthetic DMARDs in-utero for maternal treatment of rheumatic disease: a systematic review of the literature. Rheumatology 2021;60(Supplement_1): keab247.197.
10. De Cock D, Brants L, Soenen I, et al. A systematic review on the effect of DMARDs on fertility in rheumatoid arthritis. Semin Arthritis Rheum 2020;50(5):873–878.
11. Bharti B, Lee SJ, Lindsay SP, et al. Disease severity and pregnancy outcomes in women with rheumatoid arthritis: results from the Organization of Teratology Information Specialists Autoimmune Diseases in Pregnancy Project. J Rheumatol 2015;42(8):1376–1382.
12. Peters MD, Godfrey C, McInerney P, et al. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, eds. JBI Manual for Evidence Synthesis. JBI; 2024.
13. Götestam Skorpen C, Hoeltzenbein M, Tincani A, et al. The EULAR points to consider for use of antirheumatic drugs before pregnancy, and during pregnancy and lactation. Ann Rheum Dis 2016;75(5): 795–810.
14. Russell MD, Dey M, Flint J, et al; BSR Standards, Audit and Guidelines Working Group. British Society for Rheumatology guideline on prescribing drugs in pregnancy and breastfeeding: immunomodulatory

anti-rheumatic drugs and corticosteroids. Rheumatology Oxford 2023;62(4):e48–e88.
15. García-Vicuña R, Martín-Martínez MA, Gonzalez-Crespo MR, et al. Recommendations by the Spanish Society of Rheumatology for the management of patients diagnosed with rheumatoid arthritis who cannot be treated with methotrexate. Reumatol Clin 2017;13(3): 127–138. In Spanish.
16. Fernandes BM, Guimarães F, Almeida DE, et al. Portuguese recommendations for the use of biological and targeted synthetic disease-modifying antirheumatic drugs in patients with rheumatoid arthritis - 2020 update. ARP Rheumatol 2022;1(1):63–82.
17. Duarte AC, Santos-Faria D, Maria João Gonçalves MJ, et al. Portuguese recommendations for the use of methotrexate in rheumatic diseases - 2016 update. Acta Reumatol Port 2017;42(2):127–140.
18. Fiehn C, Ness T, Weseloh C, et al; DGRh Kommission Pharmakotherapie. Safety management in treatment with antimalarials in rheumatology. Interdisciplinary recommendations on the basis of a systematic literature review. Z Rheumatol 2021;80(Suppl 1):1–9.
19. Kleinmann J, Tubach F, Le Guern V, et al. International and multidisciplinary expert recommendations for the use of biologics in systemic lupus erythematosus. Autoimmun Rev 2017;16(6):650–657.
20. Alhajeri H, Abutiban F, Al-Adsani W, et al. Kuwait association of rheumatology 2018 treatment recommendations for patients with rheumatoid arthritis. Rheumatol Int 2019;39(9):1483–1497.
21. Louthrenoo W, Kasitanon N, Katchamart W, et al. 2016 updated Thai Rheumatism Association Recommendations for the use of biologic and targeted synthetic disease-modifying anti-rheumatic drugs in patients with rheumatoid arthritis. Int J Rheum Dis 2017;20(9):1166–1184.
22. Kameda H, Fujii T, Nakajima A, et al; Japan College of Rheumatology subcommittee on the guideline for the use of methotrexate in patients with rheumatoid arthritis. Japan College of Rheumatology guideline for the use of methotrexate in patients with rheumatoid arthritis. Mod Rheumatol 2019;29(1):31–40.
23. Lau CS, Chia F, Dans L, et al. 2018 update of the APLAR recommendations for treatment of rheumatoid arthritis. Int J Rheum Dis 2019; 22(3):357–375.
24. Ho CTK, Mok CC, Cheung TT, et al; Hong Kong Society of Rheumatology. Management of rheumatoid arthritis: 2019 updated consensus recommendations from the Hong Kong Society of Rheumatology. Clin Rheumatol 2019;38(12):3331–3350.
25. Sammaritano LR, Bermas BL, Chakravarty EE, et al. 2020 American College of Rheumatology guideline for the Management of Reproductive Health in Rheumatic and Musculoskeletal Diseases. Arthritis Care Res (Hoboken) 2020;72(4):461–488.
26. Cardiel MH, Carrillo S, Pérez M, et al. Update of the Mexican College of Rheumatology guidelines for the Pharmacological Treatment of Rheumatoid Arthritis, 2018. Reumatol Clin (Engl Ed) 2018. 2021; 17(4):215–228.
27. Alpízar-Campos R, Díaz-Coto JF, Vega-Ortiz JM, et al. Management guidelines of Rheumatoid Arthritis Consensus 2016. Acta Med Costarric 2017;59(3):103–109.
28. Saavedra Salinas MA, Barrera Cruz A, Cabral Castañeda AR, et al. Clinical practice guidelines for the management of pregnancy in women with autoimmune rheumatic diseases of the Mexican College of Rheumatology. Part II. Reumatol Clin 2015;11(5):305–315.
29. Pons-Estel BA, Bonfa E, Soriano ER, et al; Grupo Latino Americano de Estudio del Lupus (GLADEL) and Pan-American League of Associations of Rheumatology (PANLAR). First Latin American clinical practice guidelines for the treatment of systemic lupus erythematosus: Latin American Group for the Study of Lupus (GLADEL, Grupo Latino Americano de Estudio del Lupus)-Pan-American League of Associations of Rheumatology (PANLAR). Ann Rheum Dis 2018; 77(11):1549–1557.
30. Jaouad N, Oulkadi L, Bentaleb I, et al. Recommendations of the Moroccan Society of Rheumatology (SMR) for the Therapeutic Management of Spondyloarthritis (SpA) including Psoriatic Arthritis (PsA). Mediterr J Rheumatol 2023;34(2):139–151.
31. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021;11(4): 116–129.
32. Brouwers MC, Kho ME, Browman GP, et al; AGREE Next Steps Consortium. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ 2010;182(18):E839–E842.
33. Balshem H, Helfand M, Schünemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64(4): 401–406.
34. OCEBM Levels of Evidence. Centre for Evidence-Based Medicine. Accessed November 17, 2024. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
35. Shekelle PG, Wachter RM, Pronovost PJ, et al. Making health care safer II: an updated critical analysis of the evidence for patient safety practices. Evid Rep Technol Assess (Full Rep) 2013;(211):1–945.
36. Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. BMJ 2001;323(7308):334–336.
37. SIGN 50: a guideline developer. Healthcare Improvement Scotland. Accessed November 17, 2024. https://www.sign.ac.uk/using-our-guidelines/methodology/sign-50-a-guideline-developers-handbook/
38. Förger F, Pluma Sanjurjo A, Rüegg L, et al; On Behalf of: EULAR Task force in as Working Group: "EULAR task force QoC 012: Update of the EULAR Points to Consider for Use of Antirheumatic Drugs in Reproduction, Pregnancy and Lactation." AB1439 update of the EULAR Points to Consider for Use of Antirheumatic Drugs in Reproduction, Pregnancy and Lactation. Ann Rheum Dis. 2024;83:2075–2076.
39. Ehrlich S, Long M, Kane S, et al. Placental transfer of upadacitinib among pregnant women with inflammatory bowel disease. Inflamm Bowel Dis 2025;31(5);1489–1491.
40. Mahadevan U. Pregnancy and IBD: preliminary results from the Helmsley Global Consensus Grade Conference. Presented at: Digestive Disease Week; May 18, 2024; Washington, DC.
41. Grosen A, Bellaguarda E, Nersting J, et al. Low-dose methotrexate therapy does not affect semen parameters and sperm DNA. Inflamm Bowel Dis 2022;28(7):1012–1018.
42. De Lorenzo R, Ramirez GA, Punzo D, et al. Neonatal outcomes of children born to mothers on biological agents during pregnancy: state of the art and perspectives. Pharmacol Res 2020;152:104583.
43. Cataudella S, Lampis J, Agus M, et al. A pilot study of the relationship between pregnancy and autoimmune disease: exploring the mother's psychological process. Front Psychol 2019;10:1961.
44. Andreoli L, Gerardi MC, Gerosa M, et al; P-RHEUM.it Investigators of the Italian Society for Rheumatology (SIR). Management of pregnancy in autoimmune rheumatic diseases: maternal disease course, gestational and neonatal outcomes and use of medications in the prospectiveItalian P-RHEUM.it study. RMD Open 2024;10(2):e004091.
45. Birru Talabi M, Eudy AM, Jayasundara M, et al. Tough choices: exploring medication decision-making during pregnancy and lactation among women with inflammatory arthritis. ACR Open Rheumatol 2021;3(7):475–483.
46. Peláez-Ballestas I, Taylor CMS, Barbosa LP, et al. Pos1195-Hpr reproductive health in autoimmune rheumatic diseases from the patient's perspective: qualitative study. Ann Rheum Dis 2023;82-(suppl 1):930.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 77, No. 10, October 2025, pp 1282–1284

LETTER

DOI 10.1002/acr.25534

A call for individualized approaches to tuberculosis screening practices among new users of biologic or targeted synthetic disease-modifying antirheumatic drugs: comment on the article by Roberts et al

To the Editor:

After reading the study by Roberts et al[1] on latent tuberculosis (TB) screening among new users of biologic or targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARD), we would like to make the following comments. First, the study suggests increasing TB screening rates among new users of JAK inhibitors (JAKis) and interleukin-17 inhibitors (IL-17is), which were much lower than other drugs. Yet, clinical reality shows that reduced TB screening for JAKis and IL-17is users is justified. Studies indicate a low active TB infection incidence in JAKis users (≤0.25%)[2] and IL-17is users (0.46%, 95% confidence interval 0–1.06%),[3] which suggests that they do not increase the risk of reactivation of latent TB infection or latent TB infection reactivation risk. These data support further discussion of the need for TB screening with JAKis and IL-17is.

Second, individualized TB screening is crucial for different disease groups. For example, patients with rheumatoid arthritis (RA) and spondyloarthritis (SpA) differ in TB risk because of varying disease characteristics, patient features, and treatment protocols. Patients with RA who are older and often receive prolonged immunosuppressive therapy and multiple medications, including hormones, face higher TB risks.[4] In contrast, patients with SpA who are younger and receive less hormone therapy have lower TB risks. Thus, we suggest a precise TB screening program tailored to the disease characteristics and specific condition of each patient.

Third, as the cost-effectiveness of TB screening correlates with TB incidence,[5] TB screening in low-risk populations has low cost-effectiveness and limited benefits because of low TB incidence and screening positivity rates. Using limited resources to screen low-risk populations may disperse resources needed for high-risk groups. Studies have shown that high-risk screening is a cost-effective strategy in low TB incidence countries.[6] To improve resource efficiency and optimize health protection, we propose that TB screening strategies should be based on the disease risk level of the target population, which can focus screening resources on high-risk groups.

In conclusion, Medicare data give us a good indication that physicians should consider the potential TB risk of medications when prescribing TB screening and should target high-risk patients instead of screening all b/tsDMARD users. This improves resource efficiency and reduce patient burden. Future clinical practice guidelines are expected to offer specific recommendations, which are beneficial to achieve precision medicine, optimize patient management, and ensure patient safety.

This work was supported by the Social Development Science and Technology Projects of Wenling (2024S00268); Beijing Bethune Public Charitable Foundation (TY149EN).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr.25534.

Yuting Wang, BMed
Guiyao Jin, BMed
Wenlong Wang, MMed
wwlshutcm@126.com
The First People's Hospital of Wenling (Wenzhou Medical University, Wenling Hospital)
Wenling, China
Fanzhang Meng, MD
Beijing University of Chinese Medicine
Beijing, China
Chen Li, MD
casio1981@163.com
Tianjin Institute of Integrative Dermatology, Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital
Tianjin, China
and Fangshan Hospital, Beijing University of Chinese Medicine, Beijing, China

1. Roberts ET, Schmajuk G, Li J, et al. Latent tuberculosis screening among new users of a biologic or targeted synthetic disease-modifying antirheumatic drug: gaps in screening overall and among janus kinase inhibitors. Arthritis Care Res (Hoboken) 2024;76(7):1037–1044.
2. Cantini F, Blandizzi C, Niccoli L, et al. Systematic review on tuberculosis risk in patients with rheumatoid arthritis receiving inhibitors of Janus kinases. Expert Opin Drug Saf 2020;19(7):861–872.
3. Torres T, Chiricozzi A, Puig L, et al. Treatment of psoriasis patients with latent tuberculosis using IL-17 and IL-23 inhibitors: a retrospective, multinational, multicentre study. Am J Clin Dermatol 2024;25(2):333–342.
4. Zhang Z, Fan W, Yang G, et al. Risk of tuberculosis in patients treated with TNF-α antagonists: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2017;7(3):e012567.
5. Mohd Hassan NZA, Razali A, Mohd Nor Sham Kunusagaran MSJ, et al. A symptomatic approach to tuberculosis screening for high-risk groups in Malaysia: cost-effectiveness and budget impact analysis. J Clin Tuberc Other Mycobact Dis 2022;29:100334.
6. Gogichadze N, Sagrera A, Vicente JÁ, et al. Cost-effectiveness of active tuberculosis screening among high-risk populations in low tuberculosis incidence countries: a systematic review, 2008 to 2023. Euro Surveill 2024;29(12):2300614.

DOI 10.1002/acr.25533

Reply

To the Editor:

We thank Wang et al for their letter regarding our study of tuberculosis (TB) screening before the administration of a biologic or targeted synthetic disease-modifying antirheumatic drug (b/tsDMARD)[1] and for giving us the opportunity to clarify and expand on some important points.

First, Wang et al suggest that treatment with interleukin-17 inhibitors (IL-17is) or JAK-inhibitors (JAKis) does not increase the risk of reactivation of latent tuberculosis infection (LTBI), and therefore screening before administration of these medication classes may be unnecessary. Indeed, a recent, large, multinational, retrospective observational study of patients with plaque psoriasis treated with an IL-17i or IL-23is documented only one case of LTBI reactivation in 405 patients treated for 32.87 ± 20.95 months.[2] This is reassuring evidence of the type we called for in our discussion. However, it is important to note that contemporary practice is to screen and prophylactically treat all patients before drug initiation, so the true risk of LTBI reactivation is difficult to assess with current studies. Regarding JAKi, a recent review found less than 0.3% of 28,099 patients who were tofacitinib-exposed, and 4,310 patients who were baricitinib-exposed developed active TB with the authors concluding that the documented cases were likely community-acquired infections rather than LTBI reactivations.[3] However, this review summarizes evidence from randomized controlled trials (RCTs), open label extensions, and national registries, and it is likely that all patients in RCTs and open label extension studies were screened before study entry, limiting inferences about LTBI reactivation. Further, there is insufficient evidence regarding TB risk associated with upadacitinib and filgotinib. Thus, we call for continued analyses of real-world evidence sources regarding the risk of LTBI reactivation across medication classes.

We see the value of this kind of real-world evidence not to obviate screening but to enhance shared clinical decision-making. For example, some have adroitly pointed out that it may make clinical sense to preferentially choose an IL-17i when the patient presents with LTBI and a comorbidity that contraindicates TB chemoprophylaxis (eg, cirrhosis).[2] However, this kind of decision-making requires LTBI screening. We agree that shared decision-making about treatment regimens and weighing individualized risks and benefits for patients are appropriate.

Second, Wang et al advocate for targeted screening of high-risk patients because it may be a more cost-effective approach to identifying cases of LTBI in nonendemic locations like the United States. This raises two important questions: How would we best target screening efforts, and what is the cost of universal screening compared with missed cases?

There are certainly important differences in the prevalence of LTBI by demographic and clinical characteristics.[4] Risk stratification tools to predict developing active TB have been developed in Canada and Europe (TSTin4d and PERISKOPE-TB), though as inputs these models use a broad category of immunosuppression and the result of LTBI screening, which limits their utility for identifying whom to screen. We could assume that Wang et al are suggesting using a risk-based approach to screening for LTBI based on demographics and that such an approach would use nativity as it is biggest demographic predictor of TB in the United States. However, use of non-US place of birth alone would miss approximately one-quarter of TB diagnoses. From 2019 to 2023, there were 41,896 TB disease diagnoses in the United States. Of these, 11,106 were among US-born persons (27%).[5] When you consider that there is indeed a cost to screen a patient but that this cost is nearly negligible compared with the cost of therapy with a biologic or the cost of treating a case of TB,[6] and that TB reactivation is a preventable, potentially life threatening infection, we believe any small increased cost of screening is worth the investment.

Finally, the Centers for Medicare & Medicaid Services has set a National Quality Strategy Goal of achieving zero preventable harm.[7] Although immunosuppression is likely not the main driver of TB rates in the United States, it is central to patient safety because of the substantially higher risk of reactivation compared with most other risk factors. Unfortunately, TB rates have been on the rise across the United States since the COVID-19 pandemic,[8] and Kansas is currently experiencing the largest TB outbreak ever recorded in the United States.[9] This suggests screening should not only be universal among patients starting b/tsDMARDs but potentially for those receiving any immunosuppressive therapy.

This work was supported by the Agency for Healthcare Research and Quality (grant R01-HS-028024) and the National Institute for Arthritis and Musculoskeletal and Skin Diseases/NIH (grant K24-AR-074534).

The data presented here were supported by the American College of Rheumatology's Rheumatology Informatics System for Effectiveness registry. The views expressed herein represent those of the authors and do not necessarily represent the views of the American College of Rheumatology.

Eric T. Roberts, PhD, MPH
Eric.roberts@ucsf.edu
Matthew Murrill, MD, PhD
University of California, San Francisco
San Francisco, California
Gabriela Schmajuk, MD, MSc
University of California, San Francisco
Philip R. Lee Institute for Health Policy Research
and San Francisco Veterans Affairs Medical Center
San Francisco, California
Jinoos Yazdany, MD, MPH
University of California, San Francisco
and Zuckerberg San Francisco General Hospital
San Francisco, California

1. Roberts ET, Schmajuk G, Li J, et al. Latent tuberculosis screening among new users of a biologic or targeted synthetic disease-modifying antirheumatic drug: gaps in screening overall and among janus kinase inhibitors. Arthritis Care Res (Hoboken) 2024;76(7):1037–1044.
2. Torres T, Chiricozzi A, Puig L, et al. Treatment of psoriasis patients with latent tuberculosis using IL-17 and IL-23 inhibitors: a retrospective, multinational, multicentre study. Am J Clin Dermatol 2024;25(2): 333–342.
3. Cantini F, Blandizzi C, Niccoli L, et al. Systematic review on tuberculosis risk in patients with rheumatoid arthritis receiving inhibitors of Janus Kinases. Expert Opin Drug Saf 2020;19(7):861–872.
4. Mirzazadeh A, Kahn JG, Haddad MB, et al. State-level prevalence estimates of latent tuberculosis infection in the United States by medical risk factors, demographic characteristics and nativity. PLoS One 2021;16(4):e0249012.
5. Centers for Disease Control and Prevention. CDC Wonder. Centers for Disease Control and Prevention. Accessed March 17, 2025. https://wonder.cdc.gov/
6. Castro KG, Marks SM, Chen MP, et al. Estimating tuberculosis cases and their economic costs averted in the United States over the past two decades. Int J Tuberc Lung Dis 2016;20(7):926–933.
7. Centers for Medicare & Medicaid Services. CMS national quality strategy. Accessed January 10, 2024. https://www.cms.gov/medicare/quality/meaningful-measures-initiative/cms-quality-strategy
8. Centers for Disease Control and Prevention. Reported tuberculosis in the United States, 2023. Accessed March 5, 2025. https://www.cdc.gov/tb-surveillance-report-2023/commentary/index.html
9. Alatidd J. Kansas tuberculosis outbreak is now America's largest in recorded history. The Topeka-Capital Journal. January 24, 2025. Accessed March 5, 2025. https://www.cjonline.com/story/news/politics/government/2025/01/24/kansas-tuberculosis-outbreak-is-largest-in-recorded-history-in-u-s/77881467007/